공학박사 학위논문

A Data-based Downtime Forecasting Model in a Harbor: Application to Pohang New Port

하역중단 예보를 위한 자료기반 모델:
포항신항에 적용

2017 년 8 월

서울대학교 대학원
건설환경공학부

오 정 은

A Data-based Downtime Forecasting Model in a Harbor: Application to Pohang New Port

by

Jung-Eun Oh

Advisor: Kyung Duck Suh

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Department of Civil and Environmental Engineering

The Graduate School

SEOUL NATIONAL UNIVERSITY

August 2017

Abstract

A Data-based Downtime Forecasting Model in a Harbor: Application to Pohang New Port

Jung-Eun Oh
Department of Civil and Environmental Engineering
The Graduate School
Seoul National University

Unexpected downtime due to harbor agitation has been a serious problem in Pohang New Port. Such downtime occurrences have not well been predicted by field wave observation or numerical modeling around the port. Hence it is required to make an effort to enhance understanding intrinsic causes of downtime and to develop an efficient forecasting model that can be incorporated into daily port operation procedures.

Considering inefficiency of predicting downtime by conventional wave simulation models, a data-based model is developed in this study based on extensive wave observed data over years and corresponding downtime records available at multiple locations inside and outside the port. The main structure of the downtime forecasting model consists of Neural Network (NN) that predict wave parameters inside and outside the port from simulated wave data at outside the port and a classification model that predicts downtime occurrences based on information about

wave field produced by the NN. The overall predictive performance of the model was good, showing more than 80% of correct identification of the downtime occurrences on average.

In addition, spatio-temporal changes in spectral energies of various wave components during downtime events were examined by using Hilbert-Huang Transform (HHT) analysis, the details of which have never been studied so far. By conducting this analysis, a better understanding was obtained about the influence of each of gravity wave (GW), infragravity wave (IGW), and natural oscillation period (NOP) on the downtime occurrences inside the port. It also significantly contributed to check and improve the quality of the manually-written downtime records, which consequently enhance the eventual performance of the forecasting model.

Keywords: Downtime, Classification, Neural Networks, Hilbert-Huang Transform, Pohang New Port

Student Number: 2005-31040

Table of Contents

Abstract i

Table of Contents iii

List of Tables vi

List of Figures vii

List of Symbols xiv

CHAPTER 1. INTRODUCTION 1

1.1 Background and Motivation of the Study 1

1.2 Research Objectives 3

1.3 Structure of the Thesis 4

CHAPTER 2. LITERATURE REVIEW 6

2.1 Agitation in Harbor 6

2.1.1 Generation of Waves by Wind 6

2.1.2 Infragravity Waves 9

2.1.3 Oscillations in Harbor 12

2.2 Data-based Modeling Method 15

2.2.1 Hilbert-Huang Transform for Data Analysis 15

2.2.2 Neural Network for Data Prediction 19

2.2.3 Classification as a Branch of Machine Learning 25

CHAPTER 3. DATA 28

3.1 Pohang New Port 28

3.2 Collection of Data 31
3.2.1 Observed Wave Data 31
3.2.2 Simulated Wave Data 35
3.2.3 Recorded Downtime Data 39
3.3 Preliminary Analysis of Wave Data 43
3.3.1 Time Series of Wave Data 43
3.3.2 Correlations Between Wave Data 46
CHAPTER 4. HILBERT-HUANG TRANSFORM ANALYSIS 49
4.1 Hilbert-Huang Transform 49
4.1.1 Hilbert Transform 49
4.1.2 Empirical Mode Decomposition 51
4.1.3 Hilbert Spectral Analysis for IMFs 54
4.2 HHT Analysis of Sea Surface Elevation 58
4.2.1 Temporal Variation of the HHT spectra 60
4.2.2 Comparison of HHT Spectra at Multiple Stations 67
4.3 Quality Control of Downtime Data by Using HHT Analysis 71
4.3.1 Necessity of Examining Quality of Downtime Data 71
4.3.2 Modification of Downtime Data 73
CHAPTER 5. WAVE PREDICTION WITH NEURAL NETWORKS 78
5.1 Conventional Approach for Wave Prediction 79
5.2 Neural Network Prediction 82
5.3 Model Selection and Ensemble of Neural Networks 88
5.3.1 Model Selection Strategy 88

5.3.2 Ensemble Neural Networks 95
5.4 Discussion of the Prediction Models 103
CHAPTER 6. CLASSIFICATION MODEL 105
6.1 Classification as a Downtime Forecasting Model 105
6.2 Use of Predicted Wave and Modified Downtime Data 108
6.3 Test of Downtime Forecasting Model 115
CHAPTER 7. CONCLUSION 130
References 134
국문초록 143

List of Tables

Table 2.1 Restoring forces of floating body at mooring *adapted from* Goda (2010). 14

Table 2.2 Comparison of Fourier, Wavelet, and HHT methods. 18

Table 3.1 Example of the downtime records 40

Table 4.1 Downtime events listed in PRMAPO (2010) 59

Table 5.1 Model selection process for H_s (W03) 93

Table 5.2 Notation of ensemble methods 99

Table 5.3 Performance omparison of the different ensembles from the best selected model for each target variable 100

Table 6.1 Comparison of confusion matrices of model accuracy from the the four classification models with three different algorithms 113

Table 6.2 Confusion matrices of the classification results 120

List of Figures

Figure 1.1 General overview of this study 5

Figure 2.1 Classification of the relative energy spectrum of ocean waves (Munk, 1950, as cited and figured in Wikipedia "wind waves") 8

Figure 2.2 Sketch of wave envelope and bound long wave 11

Figure 2.3 Six modes of ship motions *adapted from* Goda (2010) 14

Figure 2.4 Discription of an artificial neuron 21

Figure 2.5 Division and methods of machine learning 26

Figure 3.1 Maps of Youngil Bay and Pohang New Port (source: Google Earth) ... 30

Figure 3.2 Locations of the four wave monitoring stations and three reference points in the wave simulation model 32

Figure 3.3 Map of berths in Pohang New Port with the three wave stations (*Adapted from* PRAMAPO, 2010): P1 to P8 denote the location of pier, and detailed quay numbers are described at each place. 33

Figure 3.4 Sea bottom bathymetry of the geographical area simulated by the SWAN model 37

Figure 3.5 Measured and simulated time series of the significant wave heights outside Pohang New Port during the second half of year 2014. 38

Figure 3.6 The monthly operating rate of Pohang New Port 41

Figure 3.7 General overview of the availability and use of the data 42

Figure 3.8 Comparison of H_s (the blue line for the left side of x-axi) and H_{sl} (the red

line for the right side of x-axis) at each monitoring station: (a) at W02. (b) at W04. (c) at W03. The shaded vertical regions mean the time of downtime occurences. 44

Figure 3.9 Changes of propagated H_s and H_{sl} at monitoring stations: (a) GW H_s at three stations W02, W04 and W03. (b) IGW H_{sl} at three stations W02, W04 and W03. The shaded vertical regions mean the time of downtime occurences 45

Figure 3.10 Scatter plots of wave heights at each monitoring station: (a) Significant wave height of GW, H_s at a station versus significant wave heifht of GW, H_s at another station. (b) Significant wave height of IGW, H_{sl} at a station versus significant wave heifht of GW, H_s at another station. (c) Significant wave height of IGW, H_{sl} at a station versus significant wave heifht of IGW, H_{sl} at another station. The correlation coefficient (R) corresponding to each parameter set is described on the top of each plot. 47

Figure 3.11 Scatter plots of wave heights between the monitoring data and calcuated data: (a) Significant wave height of GW, H_s at each station versus significant wave heifht of GW, H_s. that are calculated at C01. (b) Significant wave height of IGW, H_{sl} at each station versus significant wave heifht of GW, H_s that are calculated at C01. The correlation coefficient (R) corresponding to each parameter set is described on the top of each plot. 48

Figure 4.1 The sea surface elevation (SSE) data during one day (Mar.02, 2009) in

the first panel. The decomposed IMFs from the SSE data, F1 to F12, are displayed in the other panels. The shaded vertical regions mean the time of downtime occurences during the day. 53

Figure 4.2 The time variation of HHT spectra at the four wave stations from Feb.28 to Mar.4, 2009. The upper panel of each plot is related to longer period waves the NOPs of the port. The period of the lower panel include the period band of GW and IGW. 56

Figure 4.3 Comparison of marginal HHT spectrum and Fourier spectrum (black line) at the four wave stations on Mar.2, 2009. (a) at W02 (blue line for HHT spectrum). (b) at W04 (red line for HHT spectrum). (c) at W03 (yellow line for HHT spectrum). (d) at W10 (light blue line for HHT spectrum). 57

Figure 4.4 The time variation of HHT spectra at the four wave stations from Jan.30, 2009 to Feb.02, 2009. The period band of the upper panel on each plot is related to the NOPs of the port. The period of the lower panel include the period band of GW and IGW. 63

Figure 4.5 (a), (b) The SSE data at W03 and W10 (Jan.30, 2009 ~ Feb.02, 2009). (c) the significant wave heights of GW and IGW at W03 with that of GW at C01 for the same duration. 64

Figure 4.6 Comparing marginal HHT spectra at W03 for each one-day during Jan.30, 2009 ~ Feb.02, 2009. “DT=1” denotes that downtime occurred at that day, but “DT=0” means that there was no downtime occurrence at that day. 64

Figure 4.7 The time variation of HHT spectra at the four wave stations from Dec.30, 2008 to Jan.05, 2009. The period band of the upper panel on each plot is related to the NOPs of the port. The period of the lower panel include the period band of GW and IGW. 65

Figure 4.8 (a), (b) The SSE data at W03 and W10 (Dec.30, 2008 ~ Jan.05, 2009). (c) the significant wave heights of GW and IGW at W03 with that of GW at C01 for the same duration. 66

Figure 4.9 Comparing marginal HHT spectra at W03 for each one-day during Dec.30, 2008 ~ Jan.02, 2009. “DT=0” denotes that downtime occurred at that day, but “DT=1” means that there was no downtime occurrence at that day. 66

Figure 4.10 Comparison of marginal HHT spectra at multiple stations W02~W10 on Dec.22, 2008. “DT=1” denotes downtime occurrence on the day..... 69

Figure 4.11 Comparison of marginal HHT spectra at multiple stations W02~W10 on Dec.22, 2008. “DT=0” denotes no downtime occurrence on the day. .. 69

Figure 4.12 Comparison of marginal HHT spectra at multiple stations W02~W10 on Apr.13, 2009. “DT=0” denotes no downtime occurrence on the day. .. 70

Figure 4.13 Comparison of marginal HHT spectra at multiple stations W02~W10 on Apr.14, 2009. “DT=1” denotes downtime occurrence on the day. 70

Figure 4.14 (a-1) Boxplot and (a-2) histograms of significant wave heights of GW, H_s at W03 in accordance with the data with DT=0 and DT=1. The threshold (H_s = 0.213 m) indicates the outlier of the data H_s that is coincide with DT=0. (b-1) Boxplot and (b-2) histograms of significant

wave heights of IGW, H_{sl} at W03 in accordance with the data with DT=0 and DT=1. The threshold (H_s = 0.077 m) indicates the outlier of the data H_{sl} that is coincide with DT=0. 76

Figure 4.15 An example of quality control of the original downtime data during Apr.13, 2009 ~ Apr.17, 2009. (a) the significant wave heights of GW and IGW at W03 with the original downtime sequence. (b) the time variation of HHT spectra at W03. (c) (a) with the modified downtime sequence. The green shaded area represents the time spans corresponding to the original downtime records, whereas the vertical magenta lines denote the time instances of the revised downtime records including the confirmed and modified downtime data. 77

Figure 5.1 Scatter plots of predicted significant wave height of IGW, H_{sl} inside the harbor (at W03) by using significant wave height of GW, H_s outside the harbor (at W04). 81

Figure 5.2 Conceptual diagram of wave prediction models 84

Figure 5.3 One- and two-step prediction scheme 84

Figure 5.4 Performance comparison of one-step NN model and the nonlinear regression model 85

Figure 5.5 Performance comparison of one- and two-step NN prediction model .. 86

Figure 5.6 An exemplary time series showing performance of the one-step NN model 87

Figure 5.7 Comparison of R values from different models to select a best model for the train and validation sets 94

Figure 5.8 Comparison of NRMSE values from different models to select a best model for the train and validation sets 94
Figure 5.9 Performance comparison of MN1 and LC3 models for H_{sl} (W03) 99
Figure 5.10 The most effective ensemble model for the significant wave height of GW, H_s 101
Figure 5.11 The most effective ensemble model for significant wave height of IGW, H_{sl} 102
Figure 6.1 A flow chart of downtime forecasting model for practical application 107
Figure 6.2 Scatter plots comparing the original and modified downtime data 111
Figure 6.3 An exemplary time sequence of the original (upper) and modified (lower) downtime data use for the train set 112
Figure 6.4 General diagram of the four classification models depending on predictor inputs and response targets 112
Figure 6.5 Scatter plots of the classification models with Q.D. algorithm 114
Figure 6.6 Histograms of the number of downtimes and relative percentages at the three wave stations for different groups of the vessel weight 116
Figure 6.7 Downtime sequences at the three wave stations of W03, W04, and W10. 119
Figure 6.8 Downtime sequences with respect to the weight of ships 119
Figure 6.9 Comparison of the predicted and recorded downtime sequences for the vessel weight between 1,000 and 5,000 ton 123
Figure 6.10 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight between 1,000 and 5,000 ton 124

Figure 6.11 Comparison of the predicted and recorded downtime sequences for vessel weight between 5,000 and 10,000 ton 125
Figure 6.12 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight between 5,000 and 10,000 ton 126
Figure 6.13 Comparison of the predicted and recorded downtime sequences for vessel weight heavier than 10,000 ton 127
Figure 6.14 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight heavier than 10,000 ton 128

List of Symbols

Latin Uppercase

B Vector of biases

H_s Significant wave height of gravity waves

H_{sl} Significant wave height of infragravity waves

P Cauchy principal value of the singular integral.

S Spectral density function

T_p Peak wave period

T_z Mean period of gravity waves

T_{zl} Mean period of infragravity waves

Z Vector of outputs

Latin Lowercase

f_1 Lower cut-off frequency

f_2 Upper cut-off frequency

${}^{l}f$ Transfer function associated to a neuron of layer l

${}^{l}g_j$ Aggregation function associated to the neuron j of layer l

h Water depth

m_n *n*-th moment of wave spectrum

${}^{l}w_{ij}$ Input weight for neuron *j* and layer *l*

${}^{l}z_j$ Output value of neuron *i* of layer *l*

Greek Uppercase

${}^{l}\Omega$ Weights matrix of layer *l*

Greek Lowercase

${}^{l}\beta_j$ Bias associated neuron *i* of layer *l*

θ Phase

ω Frequency

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation of the Study

Unexpected downtime due to harbor agitation may interrupt port operation and reduce throughput. Harbor agitation is mostly caused by short waves or long wave-induced resonant oscillations, which are affected by the basin layout. Harbors with complex topologies and layouts are prone to experience severe agitation and downtime problems that are very hard to predict.

To ensure harbor tranquility, it is required to consider all possible wave components that may generate undesirable agitation inside a harbor. The influence of waves on vessels and berthing operations in a harbor is described in many national and international standards (*e.g.* PIANC, 1995; ROM, 2000; MOF, 2014), wherein a certain design limit in terms of harbor operation rate is suggested. Physical or numerical modeling are conventionally conducted to assess the wave height within a harbor and calculate the operation rate, which mainly concerns short-period waves (or gravity waves) like wind waves or swells shorter than about 30 seconds (PIANC, 1995; ROM, 2000; MOF, 2014). In reality, however, long-period waves (or infragravity waves) greater than about 30 seconds might be an important factor for affecting the movement of a vessel at berth and operation rate of a harbor (Bellotti and Franco, 2011). In addition, experimental or numerical methods could show relatively poor performance in

estimating the wave height of a port having complex topography.

Data-based modeling can be an alternative for analyzing and predicting unexpected downtime problems. With this approach, analysis of both short and long waves can be achieved in a more comprehensive manner by capturing the temporal and spatial relationships among the waves and downtime data. If we have accumulated observation or simulation data, the recent development of computer capability and algorithm enables us to obtain the underlying process of a phenomenon and reasonable prediction of it instead of directly modeling the world itself.

The motivation of this study was the lack of any model or system for predicting harbor agitation and subsequent downtimes in the Pohang New Port, although the port has suffered from frequent suspension of port operation due to harbor agitation during the winter season. To cope with this problem, the extensive wave monitoring program had been carried out over a decade from 2008 together with numerical modeling with some wave models. As a result, it was clarified that propagation of relatively short-period waves such as swell or wind waves into the port is the primary reason for occurring downtimes in the harbor. However, the influence of long-period waves in Pohang New Port has not been sufficiently examined so that it cannot be ruled out from possible causes of the harbor agitation and associated downtime events. In addition, it is still not possible to predict the downtimes by using the acquired field observation data or the numerically simulated wave data. By taking account of this, the present study aims to enhance understanding of the influence of shorter- and longer-period waves on the agitation in the port and to develop a data-based model that can quantitatively predict the downtime events at the quays in the harbor.

1.2 Research Objectives

The principal research objective of this study is to suggest a data-based downtime forecasting model for Pohang New Port based on the classification method where the downtimes are taken as the target parameter. As a first step to accomplishing this objective, it is required to collect observed and simulated wave data that are available inside and outside the port. Also, manually-written downtime records need to be collected and converted into a binary-type hourly data whose values are either 0 or 1 depending on the occurrence of downtimes or not.

In order to investigate the influence of longer-period waves on agitation and associated downtime in Pohang New Port, it was determined to apply Hilbert-Huang Transform (HHT) analysis because the analysis method discloses spatio-temporal evolution of wave energy over wide period ranges during the recorded downtime occurrences. By using this approach, the correlation between downtime occurrence and energy variation of different wave component, such as gravity waves (GW), infragravity waves (IGW), and further longer waves that are related to natural oscillation period (NOP) of the port, will be elucidated. In addition, quality control of the downtime data will be carried out based on the HHT analysis.

Another research objective is to build neural network (NN) models, which predict the waves around the port based on the numerically simulated wave and wind data. The predicted waves from the NN models can be predictors for the downtime forecasting model. Comparing NN models with nonlinear regression models were conducted in advance. For a more generalized model with good performance, model selection and the method of ensemble NN were adopted.

Finally, a data-based classifier is introduced by applying several classification algorithms to the quality-controlled downtime data as a response target while the wave data predicted by the NN model as a predictor. The schematic diagram presented in Figure 1.1 shows the general overview of the research works performed in this study. If forecasted offshore wind data is available by an operational wave model, the classifier developed in this study can be effectively used for forecasting probable future downtimes at the port and incorporated into the operational forecasting model as illustrated in the figure.

1.3 Structure of the Thesis

The thesis is organized as follows. Literature reviews are presented in Chapter 2. Description of all the data used in this study and results of primary analysis of the data are given in Chapter 3. In the following two chapters, details of the HHT analysis and the NN prediction models are elucidated. Then, the classification method for suggesting a downtime forecasting model is described in Chapter 6. Finally, the conclusion of this study is given in Chapter 7.

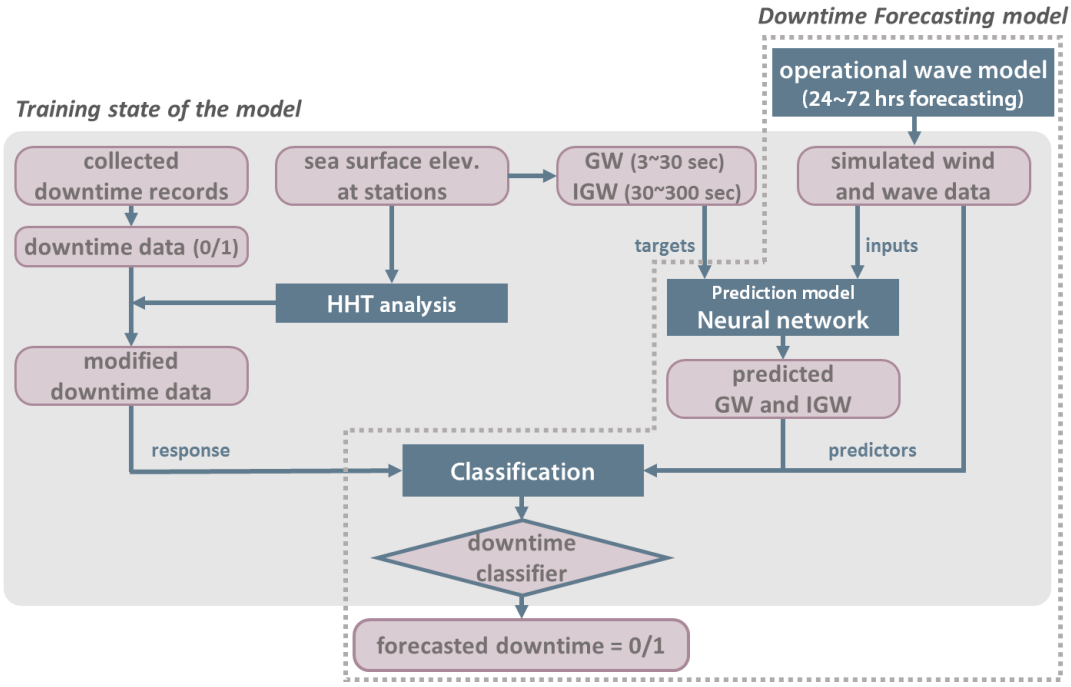


Figure 1.1 General overview of this study

CHAPTER 2. LITERATURE REVIEW

2.1 Agitation in Harbor

2.1.1 Generation of Waves by Wind

On observing the sea, both regular and random wave motions are detected continuously. Occasionally, the wind is not very strong, and the irregularities are insignificant. At certain times, large sea storms are formed by heavy wind. If the wind is strong and lasts long enough to generate gravity waves (GW), the GW do not vanish rapidly but continue to run and fetch up. The waves receive energy directly from the wind's push, and the frictional drag of the wind on the sea surface also transfer energy to the waves. Pressure differences above the waves are also a means for the waves to receive energy from the wind. If there were no energy loss, the waves would continue to grow and become indefinitely large. However, physical processes do not permit waves to increase indefinitely; sooner or later, limiting forces begin to act. The waves lose energy primarily because of the turbulent water motion. Wave breaking and viscosity of water also dissipate energy from the waves (Kinsman, 1965).

Ocean waves can be classified depending on the restoring force or disturbing force that generate the waves. In practice, waves are categorized conveniently by their period or frequency. The relative energy spectrum in Figure 2.1 roughly illustrates the energy distributed over each kind of the waves. According to this figure, the wind-generated

GW (period shorter than 30 s) contain a larger portion of the overall energy than those in any of other waves, although tidal waves contain higher energy at specific periods. This implies that the dominant force that flattens the sea surface is gravity except for tides.

GW can be divided into two states, sea wave (or wind sea) and swell. Sea wave is the wave that is under the effect of the wind that aroused it. Swell is the state in which the waves are no more under the influence of the generating wind. Sea waves have lengths shorter than those of swells and comprise steeper and coarser waves, whereas the wave directional spreading of swells is narrower than that of sea waves.

The infragravity wave (IGW), also called "long waves", "low-frequency waves" or "subharmonic gravity waves" are also restored by gravity, differently from other longer period waves. In this context, the IGW can be described as a subcategory of GW, with period of 30 s to 300 s.

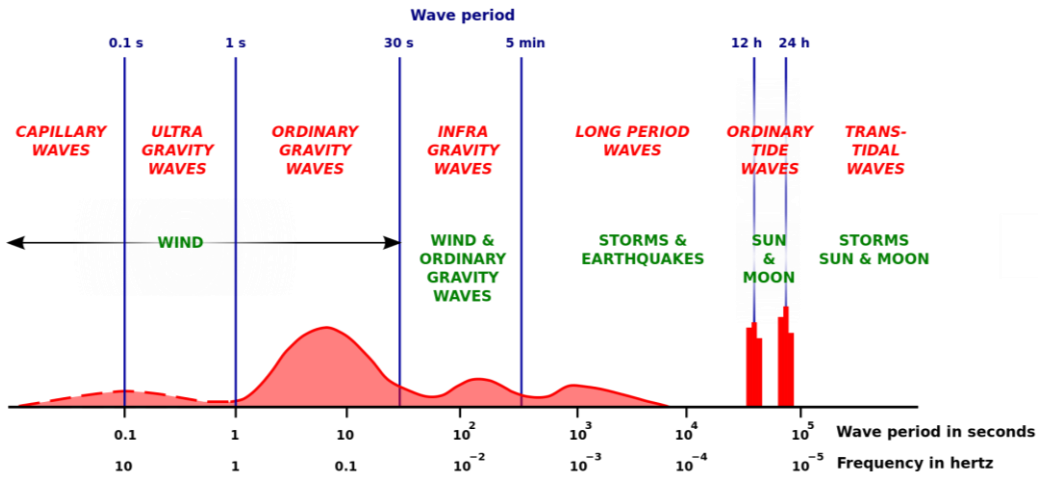


Figure 2.1 Classification of the relative energy spectrum of ocean waves (Munk, 1950, as cited and figured in Wikipedia "wind waves")

2.1.2 Infragravity Waves

IGW (infragravity wave) was first observed by Munk (1950) and Tucker (1950), respectively, in the field at La Jolla and Perranporth which are both open coasts exposed to ocean swells. They both noticed a long wave signal at a time lag relative to incoming swell wave groups. They hypothesized that the time delay was related to a reflection process near or on the coastline, but the observed time lags did not correspond well to the calculated distance by the linear wave speed. These long wave motions were subsequently confirmed by many field observations, especially in shallow water depth (Suhayda, 1974; Huntley, 1976; Wright *et al.*, 1979; Huntley *et al.*, 1981).

When two or more primary swells or sea waves, propagate to the coast with different phase speeds, the wave amplitudes sometimes add up and sometimes cancel out. This process repeats and forms a wave group with the sequence of higher and smaller wave amplitudes in deep water as depicted envelopes in Figure 2.2. In higher waves of the wave group, the mass transport tends to be negative because the fluid tends to expel. On the contrary, the fluid expelled from higher waves piles up in smaller waves. As a result, these processes bring out forced long waves underneath the wave group (see the wave drawn in the red line in Figure 2.2), and the waves are 90° out of phase for the wave group envelope. The currents exerted by radiation stress make bounded long waves beneath the wave groups of swell and wind waves (Longuet-Higgins and Stewart, 1962), and this process was verified with field and experimental observations (Elgar *et al.*, 1992; Herbers *et al.*, 1994).

The IGW propagates to the shore as bounded long waves, being highly correlated with the group of GW. In that process, the wave-wave interactions between waves may

approach resonance, the nonlinear effect takes place and enlarges the magnitude of IGW energy by the transferred energy from GW (Bowers, 1992; Mei and Agnon, 1989; Ruessink, 1998; Elgar and Guza, 1985; Herbers *et al.*, 1995).

Based on the above researches, the generation mechanism of bound IGW has been investigated to some extent, while the mechanism for the diversity that involves propagation and transformation along to the shore has not yet been fully clarified. In the nearshore region, the IGW are likely to be formed as a combination of both bound waves and free waves due to the various wave processes in the region (Herbers *et al.*, 1995; Nagai *et al.*, 1997). Meanwhile, the energy of GW is dissipated by wave breaking and some of the energy is transferred to IGW (Guza and Thorton, 1982). Due to the strong nonlinearity, it is hard to estimate precisely IGW or GW on the wave field inside a harbor.

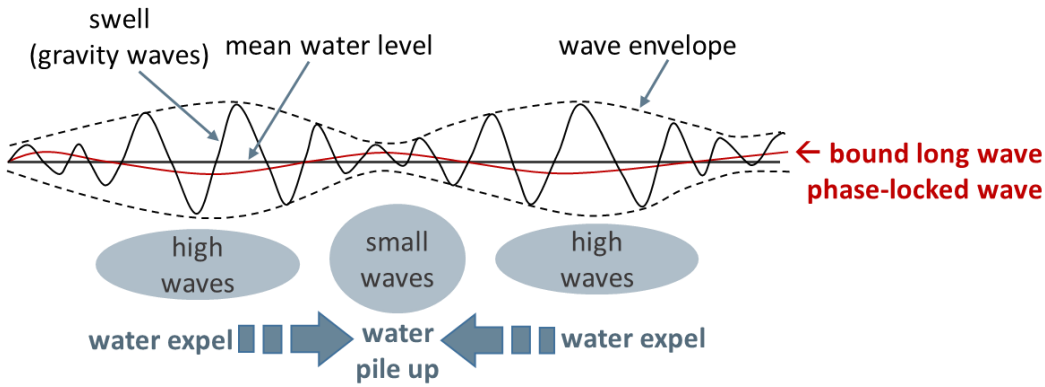


Figure 2.2 Sketch of wave envelope and bound long wave

2.1.3 Oscillations in Harbor

Harbor oscillation is the horizontal and vertical movement of sea surface elevation. On one occasion, it is just a rhythmical motion, but on the other, it takes an unexpectedly large motion. On some occasions, it persists for a long time, and on others, it decays quickly. The natural oscillation period (NOP) of a harbor as a critical factor for an appreciable oscillation usually depends on the layout of the harbor and ranges from a few seconds to a few minutes. In addition to this, the magnitude of oscillations is mainly affected by the external energy source.

Free oscillation occurs when an external force disturbs the equilibrium state of a harbor. After the oscillation subsides, the stable state is gradually or rapidly restored by gravity. In case the oscillation period is approximately equal to the NOP of the harbor, the free oscillation persists for a long time even in the absence of external forces (Rabinovich, 2009). This phenomenon is commonly designated as “seiche” or “seiching”; the components of the harbor sway back-and-forth in a rhythmic, rocking motion as gravity and friction gradually restore them to their original condition (Korgen, 1995).

On the other hand, forced oscillation begins to occur under a continuous external force. In that case, the amplitude of the oscillation depends on the proximity of the external force to the NOP of the harbor. Unlike free oscillation, forced oscillation does not persist for a long period when external forces cease. As the dominant period of external force corresponds to the NOP of the harbor, the oscillation becomes substantially amplified. This state is termed as harbor resonance. Long lasting or amplified oscillations can result in severe damages to the harbor. These oscillations have

periods that are approximately equal to the NOP of the harbor. Harbor resonance is highly undesirable; moreover, the situation will be disastrous if the oscillation period coincides with the natural period of moored vessels.

On the other hands, small harbors (5~10 m water depth and smaller than 1 km^2) can be excited by the short waves (Okihiro and Guza, 1996). Meanwhile, very long waves can affect only large harbors because their NOPs are generally longer than 10 min, which matches the very long period wave band. IGW can excite NOPs of small harbors because their NOPs are in the similar range of the infragravity wave periods (Wu and Liu, 2006).

In addition, the problem of harbor tranquility must actually be assessed from the viewpoint of ship motions (Goda, 2010). There are six modes in ship motions as shown in Figure 2.3: Vertical (heaving and pitching), longitudinal (surging and yawing), and transverse (swaying and rolling). Moored ships are affected by several restoring forces such as buoyancy, mooring lines, and fenders (see Table 2.1). Under the same wave height, the effect on ship and mode of motions are considerably changed according to wave period and mooring condition (Goda, 2010). Especially, three modes (surging, swaying, yawing) restored by forces of mooring lines tend to be affected by the long waves including IGW (López and Iglesias, 2014). Despite the changeable conditions related to ship motions, the most dangerous waves for moored ships are oscillations in the range of natural periods of horizontal ship motions. Typically, it ranges between 30 s and 5 min, which is corresponding to the wave period of IGW (van der Molen, 2006).

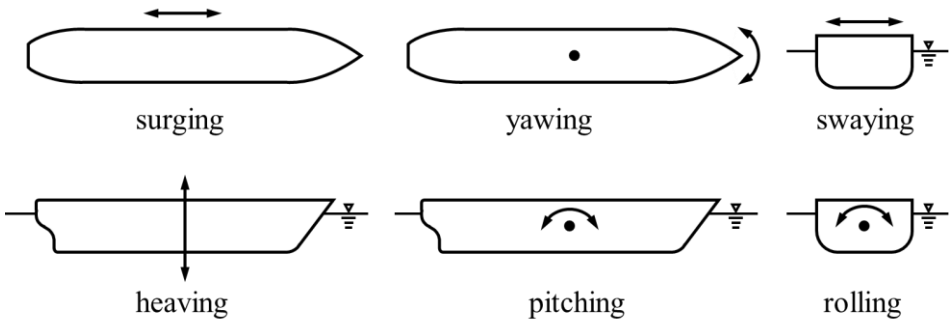


Figure 2.3 Six modes of ship motions *adapted from* Goda (2010)

Table 2.1 Restoring forces of floating body at mooring *adapted from* Goda (2010).

Mode	Buoyancy	Mooring lines	Fender
Surging	X	○	X
Yawing	X	○	△
Swaying	X	○	○
Heaving	○	X	X
Pitching	○	X	X
Rolling	○	△	X

Note: ○ active △ marginal X nonexistent

2.2 Data-based Modeling Method

There are a number of data-based modeling methods to identify or investigate the relationship among the factors in a data set. In this study, three different data-based modeling methods were employed to analyze and predict the physical phenomena associated with agitation in harbors using data are introduced here. Hilbert-Huang Transform (HHT) is very effective for presenting the spatio-temporal variation of spectral energies contained in the wave components of different periods. Neural Networks (NN), a widely used computational model for predicting target variables, can be used to predict the waves inside the harbor based on available wave data outside the harbor. Finally, classification method produces a classifier for the downtime by using the wave data predicted by the NN model.

2.2.1 Hilbert-Huang Transform for Data Analysis

A large number of data or signal processing methods have been developed so far (Poularikas, 1998). In those numerous data processing methods, Fourier transform is one of the most widely used methods. The Fourier transform is firmly established and mathematically rigorous, although it is truly valid under very general circumstances. The Fourier spectrum represents energy-frequency distributions of a given signal as a general transformation method. The perspectives of conventional Fourier methods for nonlinear mechanics arise from perturbation analysis, in which a nonlinear process is reduced to a linear system. Although limitation of such a view apparently exists for a nonlinear real world, the approach might be mathematically sound and logical.

Fourier-based methods have diverse variations, such as the spectrogram, wavelet analysis for time series, and empirical orthogonal function expansion (the principal component analysis or singular value decomposition method). Most of these methods are designed to complement the Fourier analysis, but they all fail in one way or the other (Huang *et al.*, 1998; Huang *et al.*, 1999).

Huang *et al.* (1998) point out that a reliable representation of the nonlinear and nonstationary data requires an approach that differs from the Fourier-based analysis. The new method suggested by Huang *et al.* (1998) is HHT, which is the combination of EMD and HSA. In comparison with the Fourier spectrum, the HHT spectrum can represent time–frequency (or period)–energy distribution and offer a different approach for processing time-series data.

Fourier transform divides signals into a combination of trigonometric functions as a priori basis. When a nonstationary and nonlinear natural signal is decomposed using trigonometric functions of the fixed amplitude and frequency applied to the Fourier transform, we can obtain an accurate analysis in terms of mathematics but not physics. One would inevitably face a spurious term when an irregular signal is decomposed into mathematically perfect series. It is a sound mathematical representation without useful meaning in physics. In contrast, HHT is based on an adaptive and a posteriori basis and the time–frequency–energy representation is determined through the Hilbert transform. Consequently, spurious harmonics do not require representing nonlinear waveform deformations on an a priori basis, and there is no uncertainty principle limitation on time or frequency resolution from the convolution pairs based on a priori basis.

The HHT has been applied and attempted in various fields of science, engineering, and financial analysis. In many cases, the result of HHT is much sharper than that of the

conventional methods, and it sometimes reveals true physical meanings. In testing structures, the HHT has been applied for detecting anomalies in beams and plates (Quek *et al*., 2003) and for analyzing vibration signal (Peng *et al*., 2004). HHT has also been applied to analyze earthquake signals. Huang *et al*. (2001) used the HHT spectral analysis to the earthquake data. Chen *et al*. (2004) tried to identify the natural frequencies and modal damping ratios of the Tsing Ma suspension bridge during Typhoon Victor using the HHT algorithm. In coastal engineering applications, the wave group transformation was discussed by the HHT (Veltcheva, 2002; Veltcheva and Guedes Soares, 2016a). The HHT was also adopted for analyzing the nonlinearity of abnormal waves (Veltcheva and Guedes Soares, 2016b). Hwang *et al*. (2002) compared the energy flux computation of shoaling waves by using Hilbert and Wavelet spectral analysis techniques. A brief comparison among Fourier, wavelet as a Fourier-based method, and HHT analyses are given in Table 2.2.

Table 2.2 Comparison of Fourier, Wavelet, and HHT methods.

	Fourier	Wavelet	HHT
Basis	a priori	a priori	a posteriori
Frequency	convolution over global domain, uncertainty	convolution over global domain, uncertainty	differentiation over local domain, certainty
Presentation of energy	frequency space	time-frequency space	time-frequency space
Nonlinearity	no	no	yes
Nonstationary	no	yes	yes
Feature extraction	no	discrete: no continuous: yes	yes
Theoretical base	complete mathematical theory	complete mathematical theory	empirical

2.2.2 Neural Network for Data Prediction

The NN, also called ANN (artificial neural network), is a learning algorithm that is inspired by the structure and functional aspects of biological neural networks. NN is a nonlinear statistical data modeling tool. The NN as an adaptive learning can make a generalization from previous data or extract the main features from the data. Notwithstanding, the neural network is not a novel method any longer, but it is still adaptable and extensible method.

Some natural phenomena involving ambiguous interrelationships among physical parameters have been effectively modeled and predicted by using NN. The method has been extensively applied to solve various problems related to coastal and ocean engineering fields (Kankal and Yüksek, 2012). Plenty of NN studies related to coastal and ocean areas are reviewed in Deo (2010). NN have been used for time-series forecasting as well. Recently, some hybrid approaches with NN and decomposition methods are developed and applied in time series prediction for coastal engineering field (Deka and Prahlada, 2012; Dixit and Londhe, 2016; Shahabi *et al*. 2017).

For good performance of NN, it requires appropriate settings of NN model, such as selection of NN architectures and algorithms. Various model criteria can be used for the model selection strategy to find an optimal NN model for each target data (Egrioglu *et al*., 2008). In addition, it is worth developing an ensemble of various NNs to obtain more generalized results.

2.2.2.1 Description of Neural Network

As for the NN architecture, NN consists of three kinds of layers, i.e. the input layer, hidden layer, and output layer. We have one input and one output layers with a few number of hidden layers ranges between one and three. In feedforward network, the source from the nodes at each layer propagates through the network to neurons at the next layer in a forward direction. The unit of an artificial neuron is illustrated in Figure. 5.1 for the case of *j*-th neuron at layer *l*.

The aggregation function is given by

$$ {}^{l}g_j = \sum_{i=1}^{n} {}^{l-1}z_i \; {}^{l}w_{ij} + {}^{l}\beta_j \tag{2.1} $$

where *n* is the number of inputs to neuron *j*.

Several types of functions can be applied to the transfer function as follows.
The linear transfer function:

$$ {}^{l}f\left({}^{l}g_j\right) = {}^{l}g_j \tag{2.2} $$

The logistic sigmoid function:

$$ {}^{l}f\left({}^{l}g_j\right) = \frac{1}{1+\exp\left(-\,{}^{l}g_j\right)} \tag{2.3} $$

The hyperbolic tangent sigmoid function:

$$ {}^{l}f\left({}^{l}g_j\right) = \tanh\left({}^{l}g_j\right) = \frac{2}{1+\exp\left(-2\,{}^{l}g_j\right)} - 1 \tag{2.4} $$

In every input and hidden layers, different transfer functions could be used.

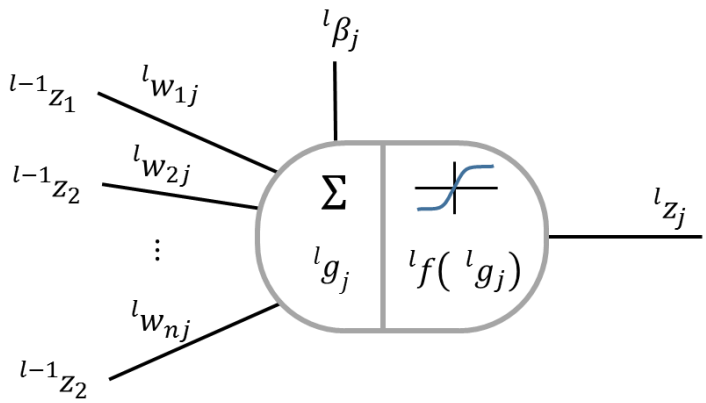


Figure 2.4 Discription of an artificial neuron

where the letters and indices are defined as follows:

$^{l-1}z_i$: the input values from neuron *i* of layer *l-1*

$^{l}z_j$: output value of neuron *i* of layer *l*

$^{l}w_{ij}$: input weight for neuron j and layer *l*, which is associated to the input $^{l-1}z_i$

$^{l}\beta_j$: bias associated neuron *i* of layer *l*

$^{l}g_j$: aggregation function associated to the neuron *j* of layer *l*

$^{l}f(\ ^{l}g_j)$: transfer function associated to the neuron *j* of layer *l*

Let ${}^{l}\Omega$ be the weights matrix of layer l, ${}^{l}B$ the corresponding vector of biases, and ${}^{l}Z$ the corresponding vector of outputs as respectively defined as following equations,

$${}^{l}\Omega = \begin{bmatrix} w_{11} & \cdots & w_{1m} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nm} \end{bmatrix} \tag{2.5}$$

$${}^{l}B = \begin{bmatrix} {}^{l}\beta_1 & \ldots & {}^{l}\beta_n \end{bmatrix} \tag{2.6}$$

$${}^{l}Z = \begin{bmatrix} {}^{l}z_1 & \ldots & {}^{l}z_n \end{bmatrix} \tag{2.7}$$

where m is the number of inputs to layer l, and n is its number of outputs.

Considering an NN with one hidden layer, an example of production $\hat{\mathrm{X}}$ can be given by equation (2.8).

$$\hat{\mathrm{X}} = {}^{2}\Omega\,{}^{1}Z + {}^{2}B = {}^{2}\Omega\{{}^{1}f({}^{1}\Omega\,W(\theta) + {}^{1}B)\} + {}^{2}B \tag{2.8}$$

Let $\epsilon(\Omega, B)$ be the error function for comparison of the estimated output $\hat{\mathrm{X}}$ with the actual target X. The mean square error function as a common error function is as follows,

$$\epsilon(\Omega, B) = E\left[(\mathrm{X} - \hat{\mathrm{X}})^{T}(\mathrm{X} - \hat{\mathrm{X}})\right] \tag{2.9}$$

The goal of the learning phase is to find the neural weight and biases that minimize $\epsilon(\Omega, B)$. To this end, the sensitivity of the outputs to each weight and bias are calculated, and some optimization methods are adopted. There are many kinds of optimization methods that update the weight and bias values; the Levenberg-Marquardt backpropagation, the conjugate gradient backpropagation, the scaled conjugate gradient

method, the Bayesian regularization backpropagation, and the gradient descent with momentum backpropagation, and so on.

The performance of NN firstly depends on proper configuration of network architecture and well-suited optimization method. In addition to this, the train set data used for network training to need to be representative of all the possible events; it is feasible to evaluate the trained network in the test set. If the performance both in the test set and train set are good enough, the result of the trained network in future data set might also be good. Thus, we need to not only select most reliable network architecture and optimization method but also pay attention to take the train and test set from the available data.

2.2.2.2 Selection and Generalization of Neural Network Model

Neural Network Model Selection

An NN model is determined by the number of hidden layers and their neurons, transfer functions, and the optimal value of the training algorithm. As these factors vary, numerous candidates can be proposed for the NN model. Selecting the most appropriate one among these is significantly important for its validity. The number of weighting parameters to be calculated in the optimization algorithm is defined by the number of hidden layers and neurons within each layer.

Moreover, if the number of weighting parameters is too small or large for the NN of a target, it would cause underfitting or overfitting issues. In other words, the complexity of the data is not well modeled with an insufficient number of parameters. Meanwhile, the excessively large number of parameters results in over-training of the NN and

subsequent poor performance; for instance, the trained network using the train data set gives irrelevant outcomes even for similar cases in the test set. Therefore, it is crucial to find a well-suited model while avoiding overfitting or underfitting.

For this end, it is necessary to select the best model after trying various configurations of the NN model based on a comparison of the model results. Different network architectures need to be considered as the first step for making a model with diverse settings. In this respect, the composition of the input parameters and the number of hidden layers and their neurons can be adjusted. In addition, different transfer functions and training algorithms can be applied. In order to compare the model performance with different configuration and to select the best, it is necessary to use various criteria and selection strategy.

Ensemble Networks

A neural network ensemble is another generalization scheme wherein many neural networks are trained for the same task. Hasen and Salamon (1990) show that the remaining generalization errors can be lessened by taking ensembles of similar neural networks (training many networks and then combining their forecasting). As mentioned above, we can select the best NN model by comparing the NN forecasts in each model, and subsequently choose the one best forecast in the best NN model based on the given criteria. However, if we build an ensemble with the NN forecasts in an appropriate manner, it will give more improved and generalized results. To create such an ensemble, it is required to select several NN forecasts for the ensemble and decide how to operate them. More details about model selection strategy and ensemble method are described in Chapter 5.

2.2.3 Classification as a Branch of Machine Learning

Machine learning is not a newly developed method (Samuel, 1959). Rather, it is being used more effectively these days owing to the advancement of computer algorithms and hardware. Mitchell (1997) provided a widely quoted definition of machine learning: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.". This saying means that doing a task based on the data has a learning effect if the performance of the task improves as the data are piled up. In this context, sufficient accumulation of the relevant data is the most important for a possible application of machine learning.

As shown in Figure 2.5, machine learning is typically categorized into supervised learning and unsupervised learning depending on whether target response labels input data or not. In this regard, unsupervised learning is useful for interpreting or grouping the data, while supervised learning is for developing a forecasting model by training the target data. Supervised learning is again divided into two methods of classification and regression depending on whether the target data is categorical or continuous. Selecting classification algorithms is very important for proper and good learning effect. Two classification algorithms used in this study are introduced here.

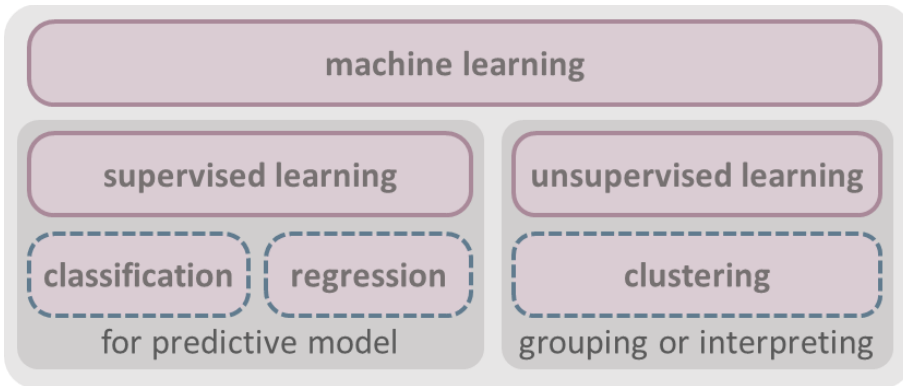


Figure 2.5 Division and methods of machine learning

k-Nearest Neighbor (kNN)

The k-nearest neighbor (kNN) classification algorithm finds a group of k objects in the training set that is nearest to the test object (Tan *et al.*, 2006). The kNN method postulates that objects close to each other are similar. To classify an object, the distance of this object to the labeled objects need to be calculated; its k-nearest neighbors are obtained, and the labels of these nearest neighbors are used to assign the label to the new object. Distance metrics, such as Euclidean, city block, cosine, and Chebychev, are used to find the nearest neighbor. The kNN classification algorithm can be refined by using the weight factor according to the distance of each k neighbors to the object (Wu *et al.*, 2008). It gives greater weighting to closer neighbors and is referred as the weighted kNN.

Discriminant Analysis

Linear discriminant analysis is from Fisher's linear discriminant and a method to find a linear combination of features that characterize two or more classes of objects (Grouven *et al.*, 1996). The method supposes different categories that are based on Gaussian distributions.; Finding the parameters for the distribution of each class is involved in training the linear discriminant analysis. When a linear discriminant analysis is not sufficiently complex for real world data, modeling for non-linear direction can be pursued for more expressive discriminant (Mika *et al.*, 1999). It is the quadratic discriminant analysis. Whether the calculated boundaries based on the distribution parameters are linear or quadratic functions, these boundaries are discriminant for the class of new objects.

CHAPTER 3. DATA

3.1 Pohang New Port

Pohang New Port is one of the largest ports on the east coast of Korea. The port is located in Youngil Bay which forms a suitable area for making a natural harbor. Nonetheless, the port mouth is slightly open to the East Sea so that tranquility inside the port is frequently deteriorated by the northeasterly waves propagated from the remote sea, especially during winter seasons.

Since 1971 when the first quay was constructed, in keeping pace with the crude steel capacity growth, Pohang New Port has been expanded into the very complex layout with eight quays in 1984 (See Figure 3.1). Worsened by the intricate topography inside the port, it has suffered from many harbor oscillation and unexpected downtime events, which lead to significant economic losses and difficulties in operating the port.

In general, downtime events are triggered by intensive swells, wind waves, or long-period waves coming from outside the harbor. To cope with such a downtime problem in Pohang New Port, several wave-monitoring programs have been conducted. Pohang Regional Maritime Affairs and Port Office (PRMAPO, 2010) performed extensive wave observation inside and outside the port and analyzed the wave conditions related to the occurrence of downtime events. A review of the recorded downtime events together with the observed waves and meteorological conditions revealed that such downtime

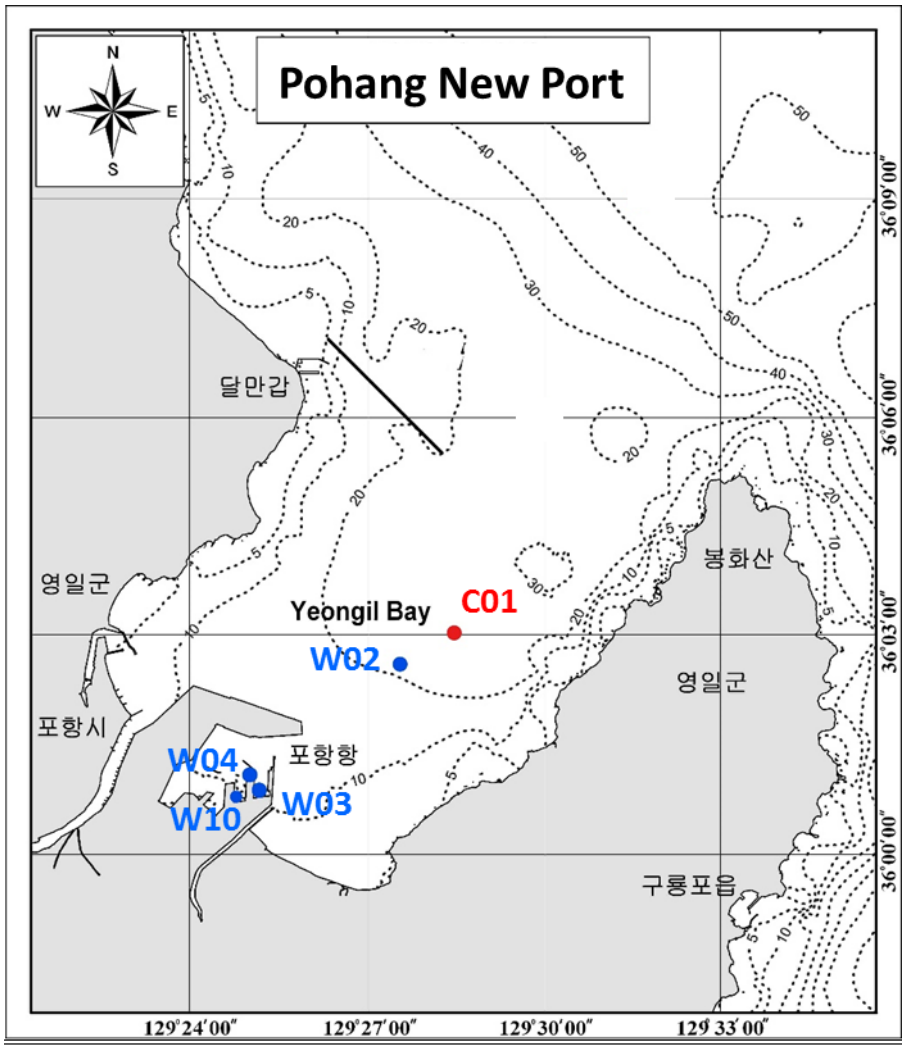
events in Pohang New Port mostly occur when incident swells or locally-generated wind waves exceed some thresholds (Jeong *et al.*, 2011; PRMAPO, 2010). Some researches, however, presented slightly different analysis results, reporting notable effects of long waves or edge waves on the downtime occurrences (Kumar *et al.*, 2014; Cho *et al.*, 2010). In this context, more extensive and detailed analysis of the observed or simulated data around the port is required to clarify the reasons for occurring agitation inside the port and consequent downtimes.

Figure 3.1 Maps of Youngil Bay and Pohang New Port (source: Google Earth)

3.2 Collection of Data

3.2.1 Observed Wave Data

This study used the observed wave data provided by PRAMAPO (2010), where extensive wave monitoring program has been performed for years at various locations inside and outside the Pohang New Port. Among the wave data available, the data at the four locations shown in Figure 3.2 were subjected to the analysis of this study. The station W02 is located outside the port, while W03, W04 and W10 stations are inside the port. In particular, downtime occurrences have most frequently occurred at W03 and W10 stations probably because they are located inside the slip surrounded by neighboring piers. In the four observation stations, wave and tide gauges (WTG) were installed to observe the wave pressure every 0.5 seconds, then the sea surface elevation (SSE) data were obtained by using the transfer function of the wave pressure. The map of Figure 3.3 shows an enlarged view of the three wave stations within the port.



Station	Water depth (m)	Latitude & longitude	
W02	21.6	36°02'15.9"N	129°27'11.3"E
W03	8.5	36°00'46.3"N	129°25'05.6"E
W04	11.0	36°01'00.1"N	129°24'58.0"E
W10	10.0	36°00'42.7"N	129°24'45.3"E
C01		36°03'00.0"N	129°28'30.0"E

Figure 3.2 Locations of the four wave monitoring stations and three reference points in the wave simulation model

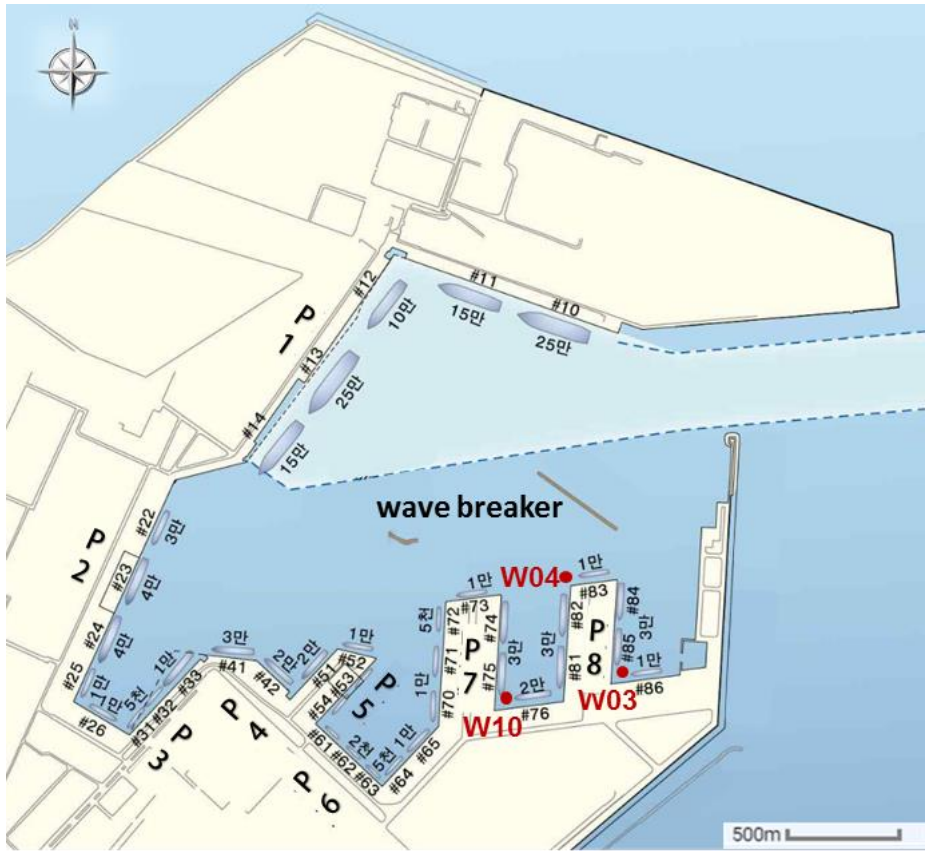


Figure 3.3 Map of berths in Pohang New Port with the three wave stations (*Adapted from* PRAMAPO, 2010): P1 to P8 denote the location of pier, and detailed quay numbers are described at each place.

Parameters of the gravity wave (GW) components were obtained by applying the spectrum analysis for the SSE data of 60 minutes. For infragravity waves (IGW), similar spectrum analysis was performed on the data of 90 minutes to obtain following parameters.

$$H_{sl} = 4\sqrt{m_0} \tag{3.1}$$

$$T_{zl} = \sqrt{\frac{m_0}{m_2}} \tag{3.2}$$

where H_{sl} and T_{zl} are the significant height and the mean period of the infragravity wave component, respectively. The symbols of m_0 and m_2 denote zeroth and second moments of the wave spectrum, respectively. The equation for calculating the moments are given by

$$m_n = \int_{f_1}^{f_2} f^n \cdot S(f) df \tag{3.3}$$

where f_1 is the lower cut-off frequency and f_2 is the upper cut-off frequency. $S(f)$ indicates a spectral density function according to the water level change for the frequency f. The frequency range of infragravity wave is determined by the values of cut-off frequencies, thus setting the lower and upper cut-off frequency in equation (3.3) is very important. Referring to the previous study (Jeong *et al.*, 2011), the upper and lower cut-off frequencies were set as 1/30 Hz and 1/300 Hz, respectively.

3.2.2 Simulated Wave Data

The simulated wave data produced by Jeong *et al*. (2016) and Jeong (2017) were also used in this study. In their researches, numerical simulation was carried out with SWAN (Simulating Waves Nearshore, version 40.91) wave model with the wind field provided by Japan Meteorological Agency (JMA) as input variables. JMA is providing three different groups of GPV (Grid Point Values) wind data, among which the Meso-Spectral Model (MSM) data were employed in this study. The spatial resolution of the wind data is 1/20° (approximately 5 km) whereas the temporal resolution is 3 hours.

Computational domain for SWAN wave modeling was built with a Cartesian coordinate covering a large-scale area of the longitude 117° to 147° East and the latitude 22° to 50° North with a grid resolution of 1/20° by 1/20°. Figure 3.4 shows the sea bottom bathymetry of the computational domain. The wave model was configured with 21 exponentially spaced frequencies ranging from 0.04 to 0.4 Hz and 72 evenly spaced directions. The SWAN model was run with the default source and sink terms for physical processes implemented in the model, and outputs were released at every 30 minutes. More detailed description of the mathematical and numerical expressions of each forcing term is presented in SWAN technical documentation (SWAN team, 2013).

The simulated wave fields were validated with the field observation data acquired by AWAC (Acoustic Wave And Current meter) that were deployed at the location shown in Figure 3.2. Figure 3.5 shows the comparison between the measured and simulated significant wave height (H_s) during the second half of the year 2014. The measured data obtained from AWAC are shown in blue line whereas the simulated results by SWAN model are shown in red line. The simulation results show reasonably good agreement

with the measurement regarding overall temporal variations of the wave height. The difference between the measurement and computation tended to be larger when the wave height was relatively larger.

To more quantitatively assess the model performance by using reliable statistical parameters, two well-applied indices evaluating model accuracy were used in the present study: the index of agreement (*Ia*) suggested by Willmott (1981) and the Pearson's correlation coefficient (*R*), which are respectively calculated as follows.

$$Ia = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2} \tag{3.4}$$

$$R = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{3.5}$$

where N is the number of data points of the time series that are considered, y_i and $\bar{y}$ denote the observed data and their mean value, and $\hat{y}_i$ and $\bar{\hat{y}}$ are the modeled data and their mean value. *Ia* or *R* value closer to 1 indicates that a greater proportion of variance is accounted for by the model. A fairly good agreement was obtained for H_s with both indices of *Ia* and *R* over 0.9 on average.

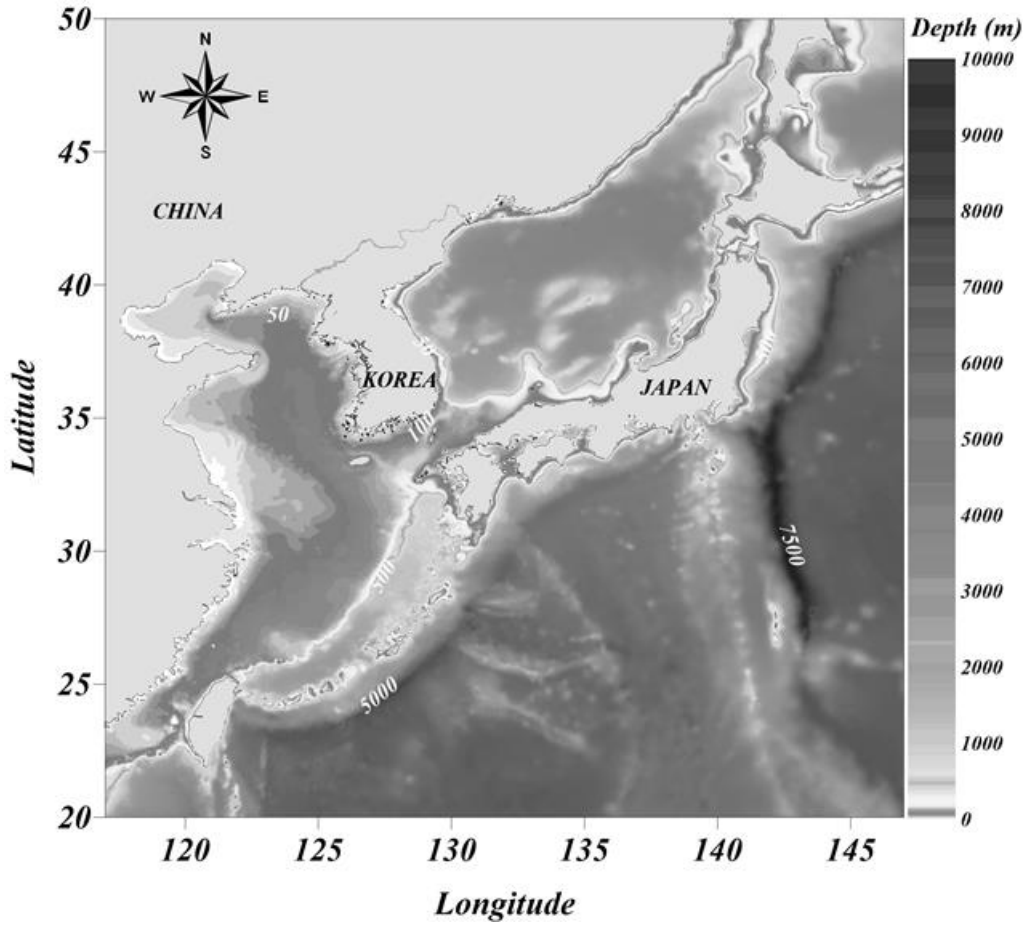


Figure 3.4 Sea bottom bathymetry of the geographical area simulated by the SWAN model

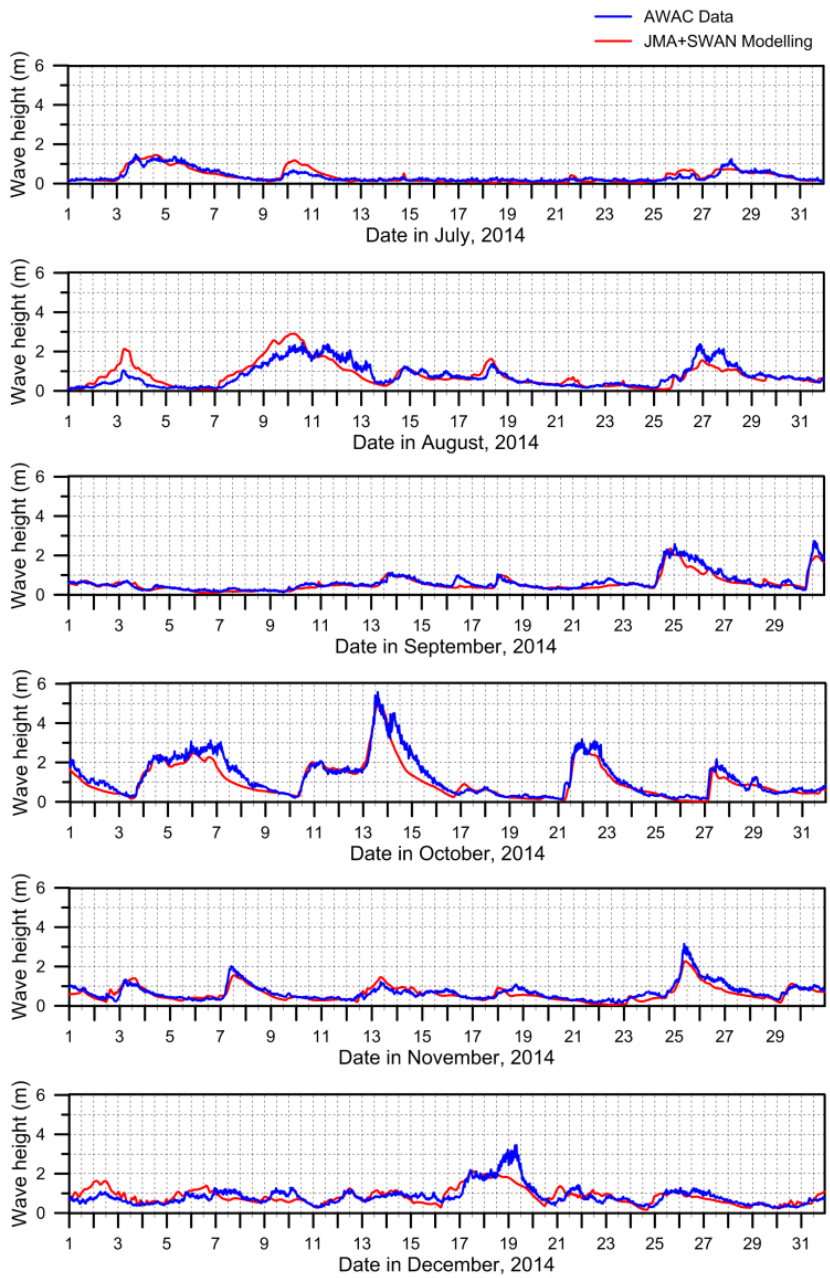


Figure 3.5 Measured and simulated time series of the significant wave heights outside Pohang New Port during the second half of year 2014.

3.2.3 Recorded Downtime Data

Records of the downtime events from Dec.18, 2008 to Dec.31, 2010 were also collected during the wave observation period. An example of downtime records is listed in Table 3.1. The downtime records consist of the hour that downtime begins and ends, quay number and weight of ship with its name.

To make hourly downtime data, any hour is told in downtime records is assigned as "1" in the downtime data, other hours are allocated as "0" in the downtime data. The yearly and monthly operating rate of Pohang New Port is shown in Table 3.2. Figure 3.6 reveals that downtime events more frequently occurred in the winter season.

In Figure 3.7, information about the availability of all the data is presented and the used period and the number of data are presented with respect to the data modeling methods. The practical use of the train, test and validation set will be described in sections 5.2, 5.3, and 6.3.

Table 3.1 Example of the downtime records

month	day	downtime hour		quay #	ship name	ship weight (ton)
		begin	end			
1	2	0	24	76A	M/KIMITSU	1680
1	2	15	24	71	SNK LADY	1691
1	2	18	24	33	LINDOS	31643
1	2	18	24	86B	WEN CHENG	2423
1	2	23	24	51	Alam Selamat	21941
1	3	0	13	24	JIA QIANG	26078
1	3	0	7	31	W.Y VENUS	1182
1	3	0	24	31	MIRAI ASTRO	1993
1	3	0	10	33	LINDOS	31643

Table 3.2 The yearly and monthly operating rate of Pohang New Port

		monthly													
		1	2	3	4	5	6	7	8	9	10	11	12	sum (hr)	operability (%)
yearly	2009	282	102	78	159	32	0	0	10	0	42	159	40	904	89.7
	2010	135	134	140	20	76	0	0	0	17	83	85	197	887	89.9
	average (hr)	208.5	118.0	109.0	89.5	54.0	0.0	0.0	5.0	8.5	62.5	122.0	118.5	895.5	
	operability (%)	72.0	82.4	85.3	87.6	92.7	100.0	100.0	99.3	98.8	91.6	83.1	84.1		**89.8**

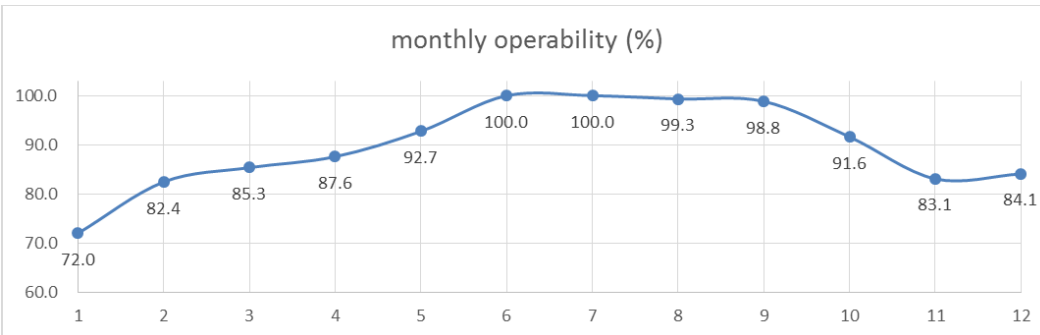


Figure 3.6 The monthly operating rate of Pohang New Port

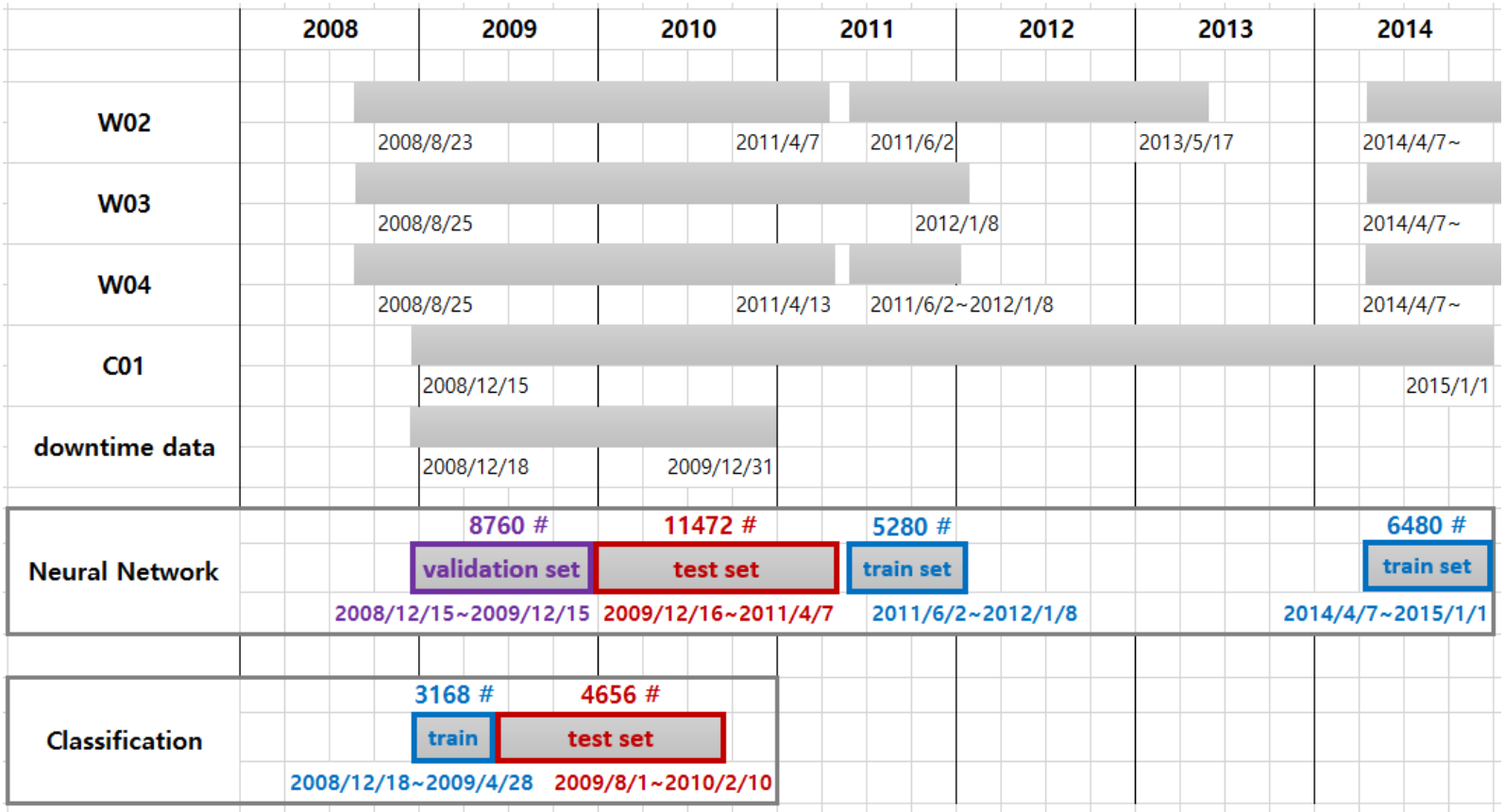


Figure 3.7 General overview of the availability and use of the data

3.3 Preliminary Analysis of Wave Data

3.3.1 Time Series of Wave Data

Figure 3.8 shows the time series of the calculated and observed significant wave height from November 2009 to February 2010. Blue and red lines in the three panels represent significant wave heights of the GW H_s and the IGW H_{sl}, respectively. Note that the magnitude of H_s is shown on the vertical axis on the left side, whereas H_{sl} on the right side. The green vertical shadings indicate the time when downtime events occurred. The figure shows that the downtime events are likely to appear when H_s and H_{sl} are relatively large. A similar tendency was also observed for the rest of the entire collected data. As seen in this observation, downtime events occurred more frequently in the winter season.

As shown in Figure 3.9, when waves propagate from the offshore to inside the harbor through the entrance of the port, the magnitudes of GW gradually decrease, and the values inside the harbor become relatively small because a lot of wave energies are dissipated at the breakwater. Although the heights of GW are largely reduced compared with those outside the harbor, they are still greater than the heights of IGW. The relationship between the occurrence of downtime events and the magnitudes of GW or IGW will be discussed later.

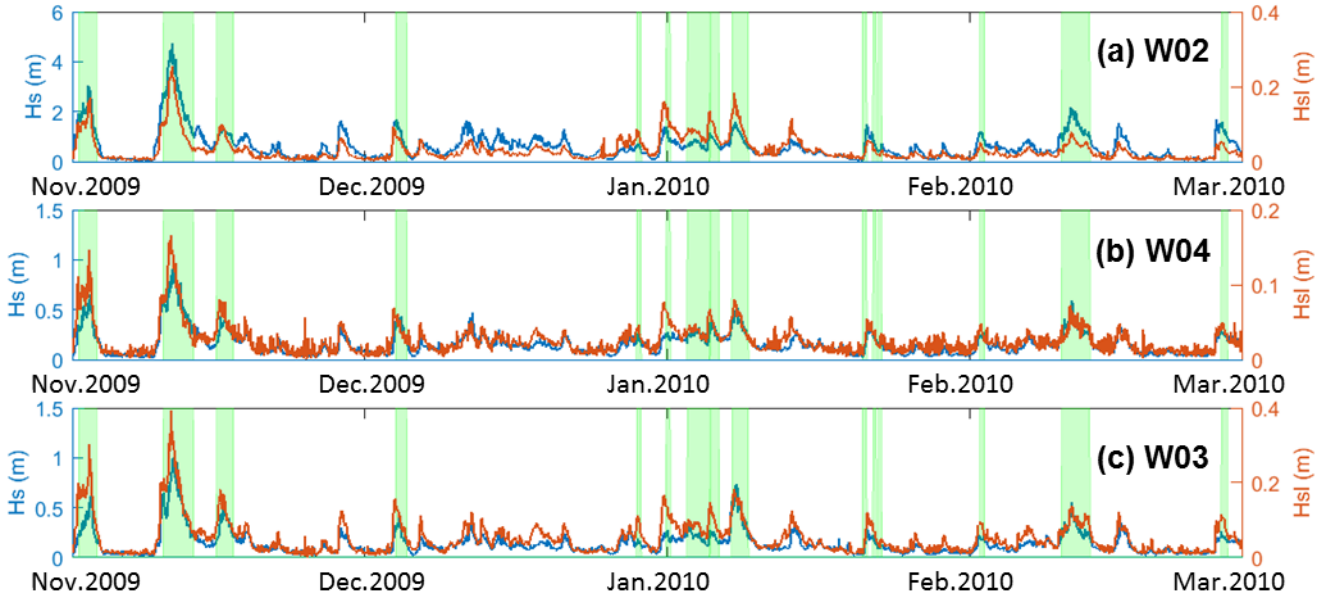


Figure 3.8 Comparison of H_s (the blue line for the left side of x-axi) and H_{sl} (the red line for the right side of x-axis) at each monitoring station: (a) at W02. (b) at W04. (c) at W03. The shaded vertical regions mean the time of downtime occurences.

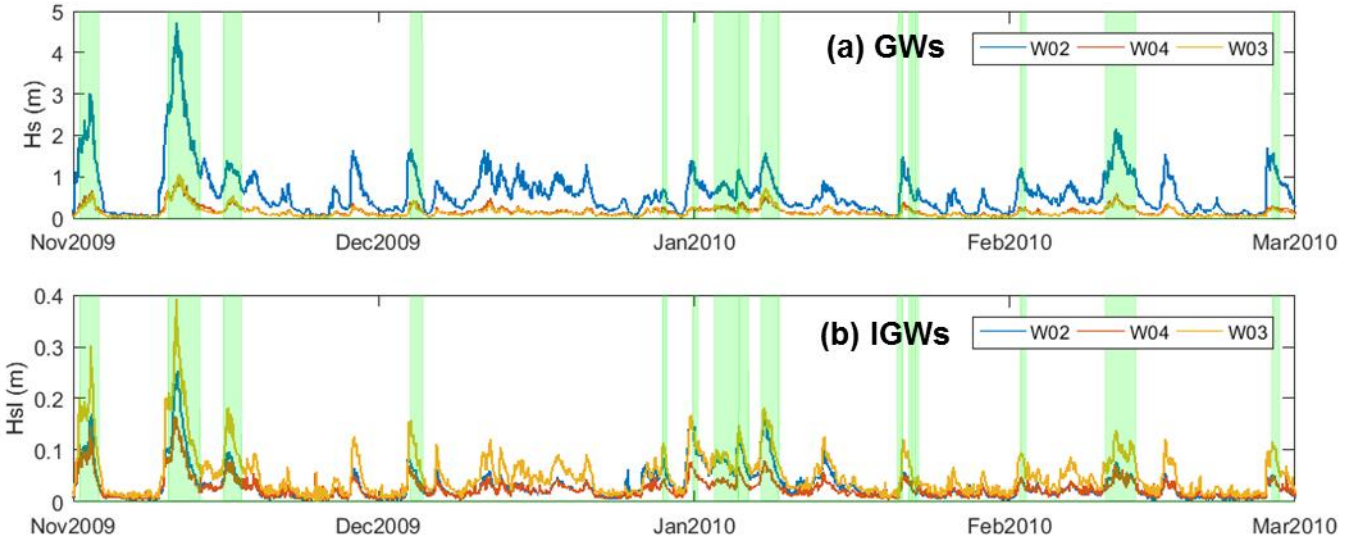


Figure 3.9 Changes of propagated H_s and H_{sl} at monitoring stations: (a) GW H_s at three stations W02, W04 and W03. (b) IGW H_{sl} at three stations W02, W04 and W03. The shaded vertical regions mean the time of downtime occurences

3.3.2 Correlations Between Wave Data

The correlation between GW and IGW was examined at the four wave monitoring stations as shown in Figure 3.10. Meanwhile, relationships between the observed and the calculated data are investigated in Figure 3.11. The scatter plots of H_{sl} and H_s among observed data reveal quadratic or close-to-linear proportionality (in Figure 3.10), but those among the calculated and observed data do not show such an apparent relationship (in Figure 3.11).

Infragravity wave heights in a nearshore region and gravity wave heights in the offshore area (Tucker, 1950). The above correlation analysis presents the possibility that some energy may be transferred from gravity wave band to infragravity wave band during the propagation of waves into the harbor. However, the relationship between the two quantities changes very rapidly during the propagation of waves from offshore to the shore through the continental shelf, surf zone, and swash zone. Such a complicated change is caused by topography and complex wave processes as well. In this context, it is necessary to understand and use the correlation between GW and IGW for reasonable prediction.

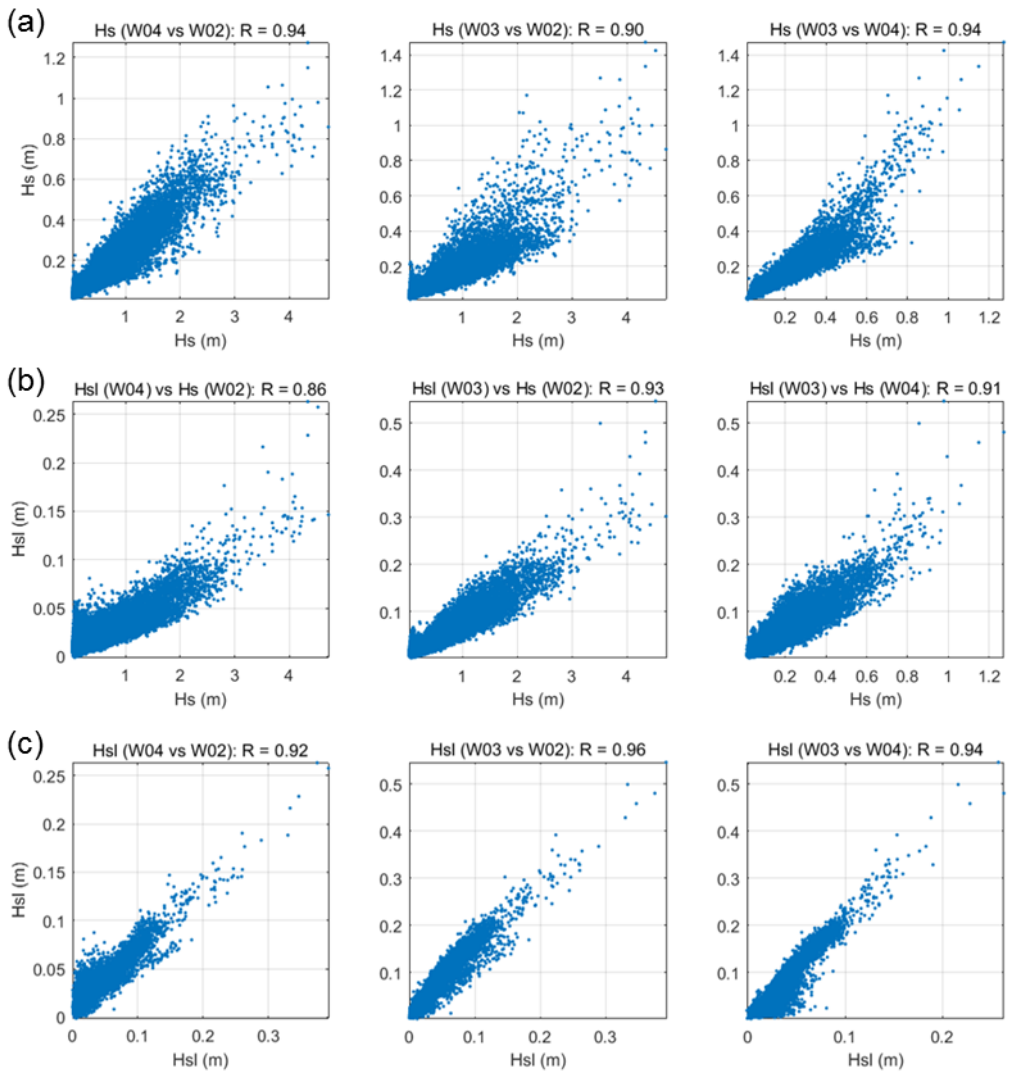


Figure 3.10 Scatter plots of wave heights at each monitoring station: (a) Significant wave height of GW, H_s at a station versus significant wave heifht of GW, H_s at another station. (b) Significant wave height of IGW, H_{sl} at a station versus significant wave heifht of GW, H_s at another station. (c) Significant wave height of IGW, H_{sl} at a station versus significant wave heifht of IGW, H_{sl} at another station. The correlation coefficient (R) corresponding to each parameter set is described on the top of each plot.

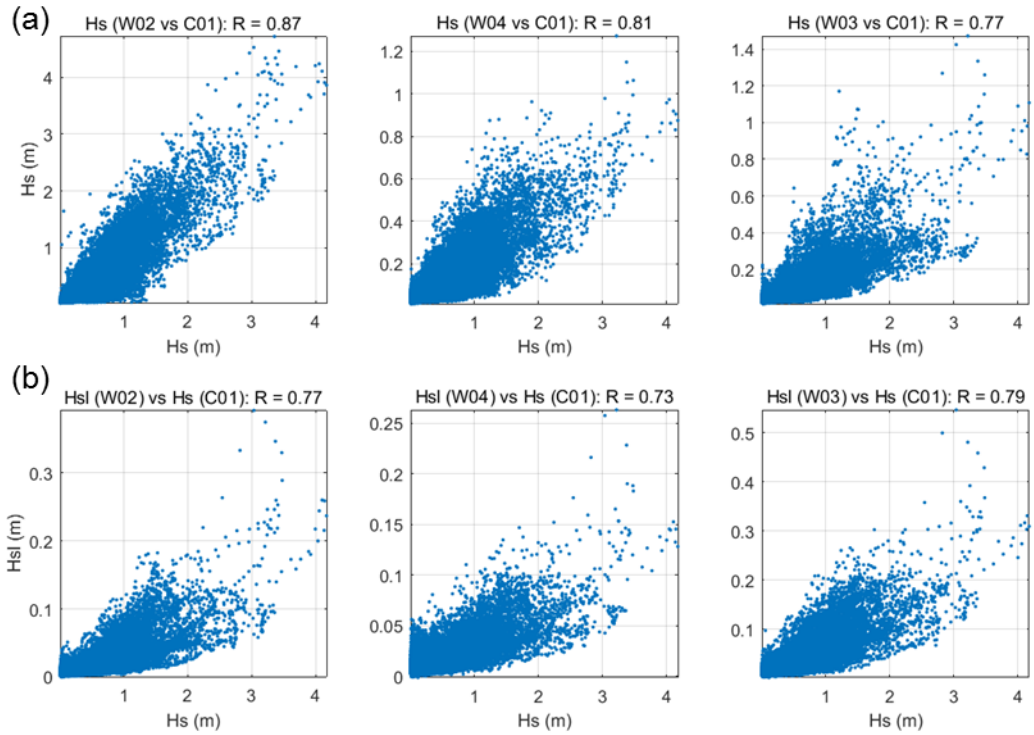


Figure 3.11 Scatter plots of wave heights between the monitoring data and calcuated data: (a) Significant wave height of GW, H_s at each station versus significant wave heifht of GW, H_s. that are calculated at C01. (b) Significant wave height of IGW, H_{sl} at each station versus significant wave heifht of GW, H_s that are calculated at C01. The correlation coefficient (R) corresponding to each parameter set is described on the top of each plot.

CHAPTER 4.

HILBERT-HUANG TRANSFORM ANALYSIS

4.1 Hilbert-Huang Transform

4.1.1 Hilbert Transform

The Hilbert-Huang Transform (HHT) was developed by Huang *et al*. (1998) to implement the Hilbert methods for practical use of real world data. HHT consists of two steps to analyze the data. The first step is to decompose the data into several functions that meet some conditions for reliable use of the Hilbert transform. The second step is to apply the Hilbert spectral analysis to the decomposed functions and construct the time-frequency (or period)-energy distribution, which is the HHT spectrum.

For an arbitrary time series $x(t)$ of L^p class, we can get the Hilbert transform $y(t)$ of $x(t)$ as follows, where P is the Cauchy principal value of the singular integral.

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{4.1}$$

By this transformation, we can obtain the analytic function,

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)} \tag{4.2}$$

From the relationship in this analytic function, we can also obtain instantaneous

amplitude $a(t)$ and instantaneous phase, $\theta(t)$.

$$a(t) = \sqrt{x^2 + y^2} \tag{4.3}$$

$$\theta(t) = \tan^{-1}\frac{y}{x} \tag{4.4}$$

Then the instantaneous frequency is derived from this phase function.

$$\omega(t) = \frac{d\theta(t)}{dt} \tag{4.5}$$

This representation suggests the new concept about frequencies determinant of phase in comparison with generally used mean frequency, $\omega = 1/T$.

The right-hand side of equation (4.2) gives an apparent time-frequency-energy expression for $x(t)$. However, the instantaneous frequency obtained using the above equations for an arbitrary function shows sudden fluctuations, and it is not necessarily physically meaningful. The Hilbert transform cannot be used as it is because of this distortion problems. Unlike widely used Fourier transform, the application of Hilbert transform is limited to some problem, such as shifting phase.

To avoid the distortion problem of the Hilbert transform for an arbitrary function, it needs a purely oscillatory function (monocomponent) with zero reference level (Huang *et al.*, 1998). The empirical mode decomposition (EMD) developed by Huang *et al.* (1996) bears some functions that meet such condition, and the Hilbert transform can obtain a meaningful representation by using these functions.

Hence, HHT uses the Hilbert transform, but its application is different from that of the Hilbert transform. HHT can be used as broadly as Fourier transform since the HHT decomposes a signal into a series of components and transforms the components by

using the Hilbert transform without the distortion problem. At the same time, those decomposed components remain nonstationary and nonlinear as they were, so it makes HHT more favorable than Fourier transform.

4.1.2 Empirical Mode Decomposition

The devised EMD for the Hilbert transforms to make the meaningful representation can divide any arbitrary function into an open number of IMFs (Intrinsic Mode Functions). Analyzing the decomposed IMF using the Hilbert transform gives the meaningful local phase change. Another advantage of EMD is to offer an adaptive method for nonlinear, nonstationary data analysis. The sifting process of EMD enables a complex data to be distributed into amplitude and frequency modulated forms of IMFs. There are two conditions for IMF as follows.

1. The number of extrema and thus the number of zero-crossings in the whole data series must be equal or differ at the most by one.
2. At any instant in time, the mean value of the envelope of the local maxima and the envelope of the local minima is zero.

Algorithm of EMD

The algorithm of EMD to decompose and derive IMFs from a signal (or time series data) $x(t)$ is as follows.

(1) Initialize: $r_0(t) = x(t),\ i = 1$

(2) Extract the ith IMF:

(a) Initialize: $h_0(t) = r_i(t),\ k = 1$

(b) Extract the local maxima and minima of $h_{k-1}(t)$

(c) Interpolate the local maxima and local minima by cubic spline to form upper and lower envelopes of $h_{k-1}(t)$

(d) Calculate the mean $m_{k-1}(t)$ of the upper and lower envelopes of $h_{k-1}(t)$

(e) Define: $h_k(t) = h_{k-1}(t) - m_{k-1}(t)$

(f) If IMF criteria are satisfied,

then set $\mathrm{IMF}_i(t) = h_k(t)$; else go to (b) with $k = k + 1$

(3) Define: $r_i(t) = r_{i-1}(t) - \mathrm{IMF}_i(t)$

(4) If $r_i(t)$ still has at least two extrema, then go to (2) with $i = i + 1$;

else the decomposition is completed and $r_i(t)$ is the "*residue*" of $x(t)$

Through the above process of EMD, the detrending effect for the data is naturally achieved. The obtained IMFs constitute adaptive bases for the Hilbert spectral analysis (HSA). The IMFs satisfy all the major mathematical requirements, convergence, completeness, orthogonality, and uniqueness empirically.

An example of IMFs from the EMD for SSE (Sea Surface Elevation), which are resampled to 2 seconds from the original 0.5 seconds interval SSE data, is shown in Figure 4.1. The first panel is the original SSE data, while the second to last panels are IMFs derived from the original one. The shaded part is the time span where the downtime recorded. It is clear that some IMFs (*e.g.* F1~F6) are relatively amplified at that time span.

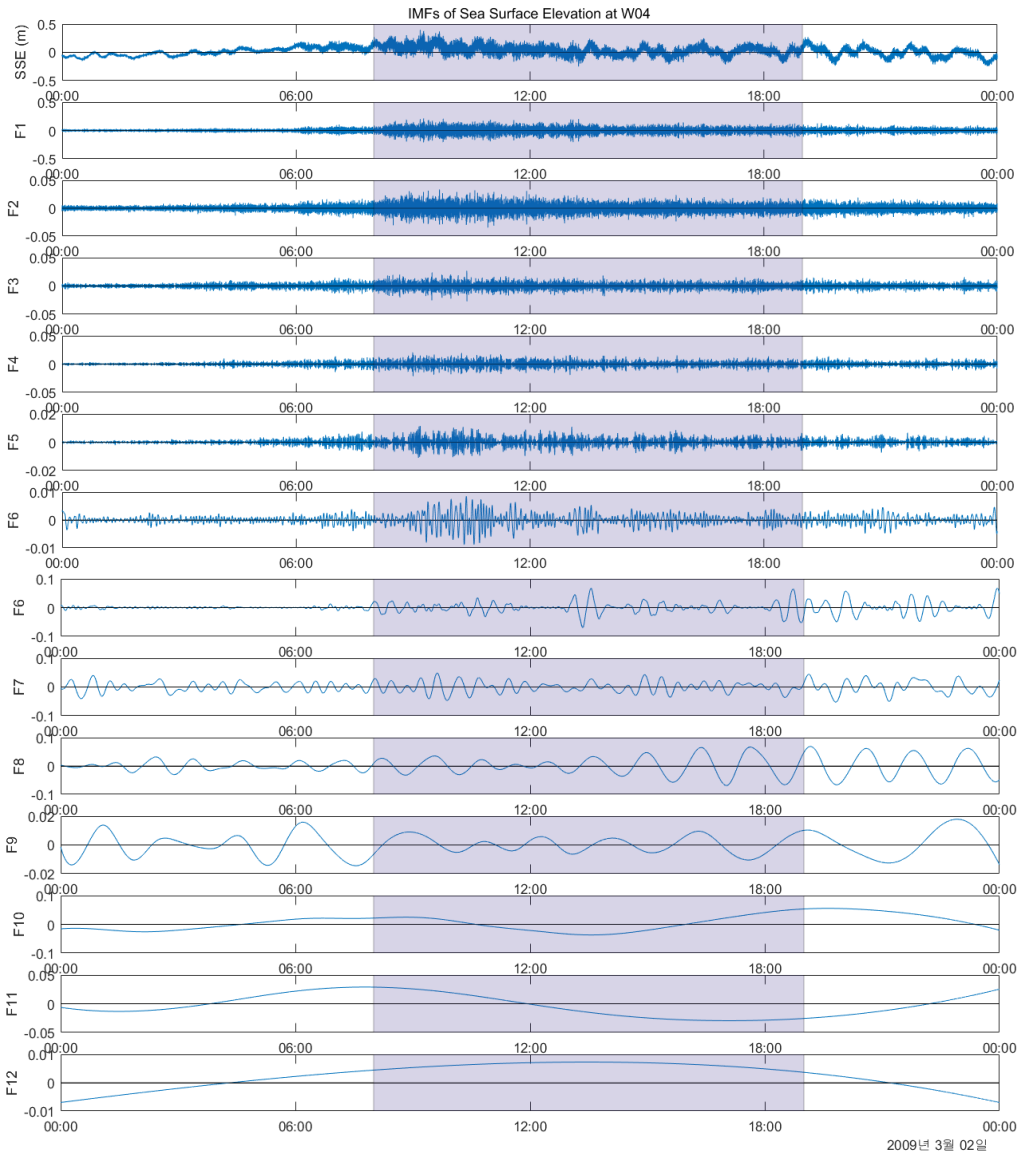


Figure 4.1 The sea surface elevation (SSE) data during one day (Mar.02, 2009) in the first panel. The decomposed IMFs from the SSE data, F1 to F12, are displayed in the other panels. The shaded vertical regions mean the time of downtime occurences during the day.

4.1.3 Hilbert Spectral Analysis for IMFs

Applying the Hilbert transform for n numbers of decomposed IMFs gives the representation of an arbitrary function $x(t)$ as follows,

$$x(t) = \mathrm{Re}\left[\sum_{j=1}^{n} a_j(t)\, e^{-i\omega_j(t)dt}\right] \tag{4.6}$$

where Re[ ] denotes the real part of terms within the bracket. $a_j(t)$ and $\omega_j(t)$ are the instantaneous amplitude and phase function of the j-th IMF. For comparison, an arbitrary function $x(t)$ can be expressed in Fourier transform as

$$x(t) = \mathrm{Re}\left[\sum_{j=1}^{\infty} a_j\, e^{-i\omega_j t}\right] \tag{4.7}$$

In the representation of Fourier transform, a_j and ω_j are constants.

From the above two equations, the Hilbert representation using IMFs can be considered as the generalized Fourier expansion. The harmonic representation of the Fourier transform using priori bases inevitably involves some spurious harmonics that are just made for mathematical representation. On the other hands, HHT using adaptive bases can demonstrate the intra-wave frequency change and modulation. Therefore, HHT can accommodate the nonlinear and nonstationary variation beyond the ability of Fourier transform.

From the application of HSA to the decomposed IMFs by using equation (4.6), the time-period-energy representation of HHT spectrum can be obtained. Figure 4.2 shows the results at W02, W03, W04 and W10, where the contour plots of HHT spectrum after applying a Gaussian filter for spatial smoothing is shown. By comparing the four diagrams in the figure, the difference in HHT spectrum depending on stations is

recognized. The upper panel shows the temporal variation of spectral energy during five days at the very long period band (5 minutes to 24 hours), which can be related to the NOP of the port. Meanwhile, the lower panel shows the temporal variation at the shorter period band (6 seconds to 5 minutes) that are associated with GW and IGW. By examining those diagrams, it is possible to understand the spatio-temporal evolution of spectral energy of various wave components corresponding to different periods.

From the HHT spectrum shown in Figure 4.2, we can obtain a marginal frequency spectrum during a specified period. Figure 4.3 shows the comparison of FFT spectrum and the marginal HHT spectrum during one day (March 2nd). The marginal HHT spectrum shows apparently different energy distributions compared to the Fourier spectrum. In those figures, the Fourier spectra show very fastly-varying noisy fluctuations over the lower period band (approximately less than 5 minutes). In contrast, the marginal HHT spectra show more identifiable patterns that are more favorable to find out meaningful spectral peaks. Although it is not fair to say that the marginal HHT spectrum shows more desirable patterns physically than the Fourier spectrum, it provides more useful analysis tool for detecting significant features enclosed in the data. In other words, the variational pattern of the marginal HHT spectrum showing discernible peaks may have a statistical meaning that can be linked to the target phenomenon, such as the downtime event.

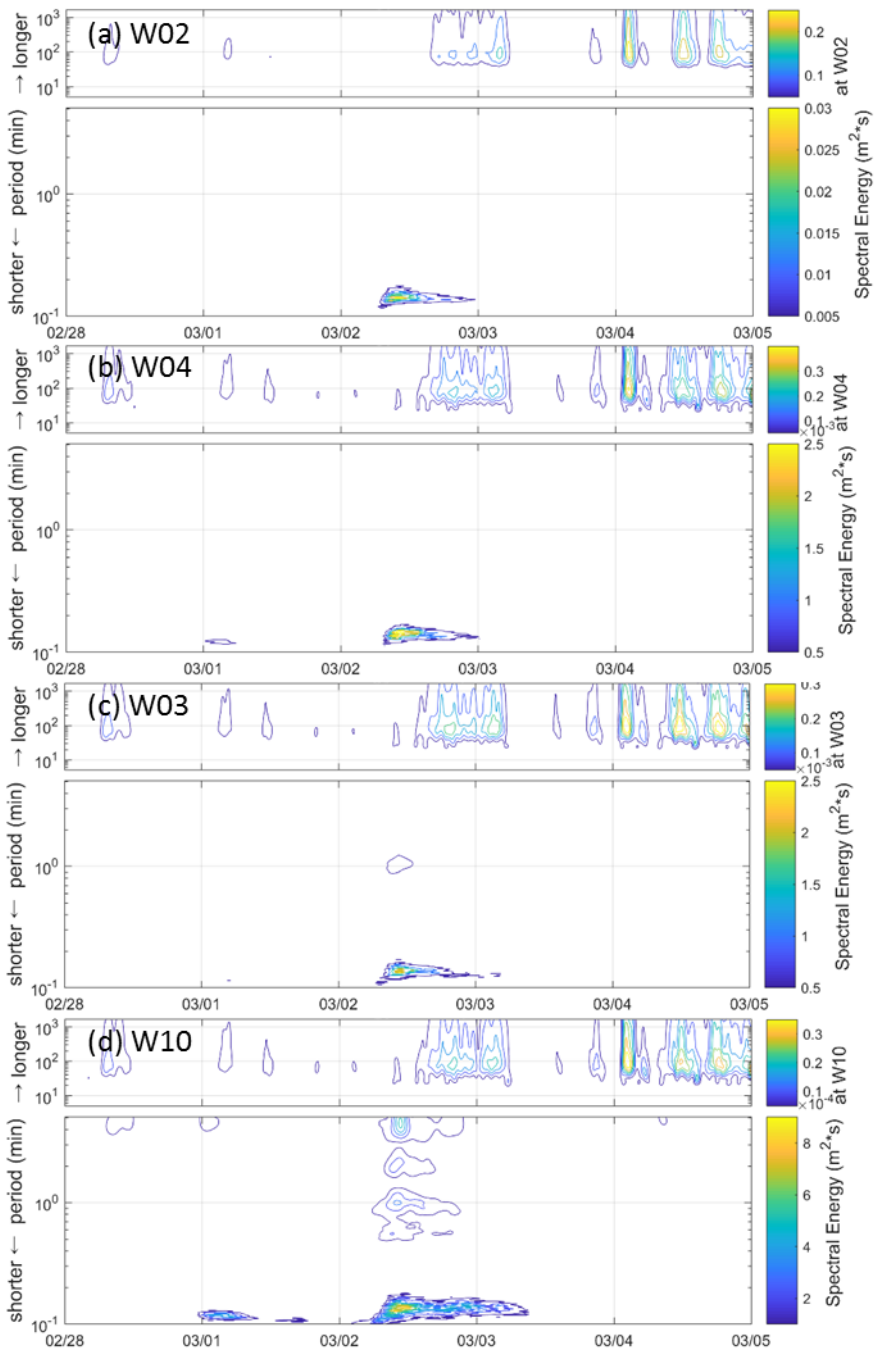


Figure 4.2 The time variation of HHT spectra at the four wave stations from Feb.28 to Mar.4, 2009. The upper panel of each plot is related to longer period waves the NOPs of the port. The period of the lower panel include the period band of GW and IGW.

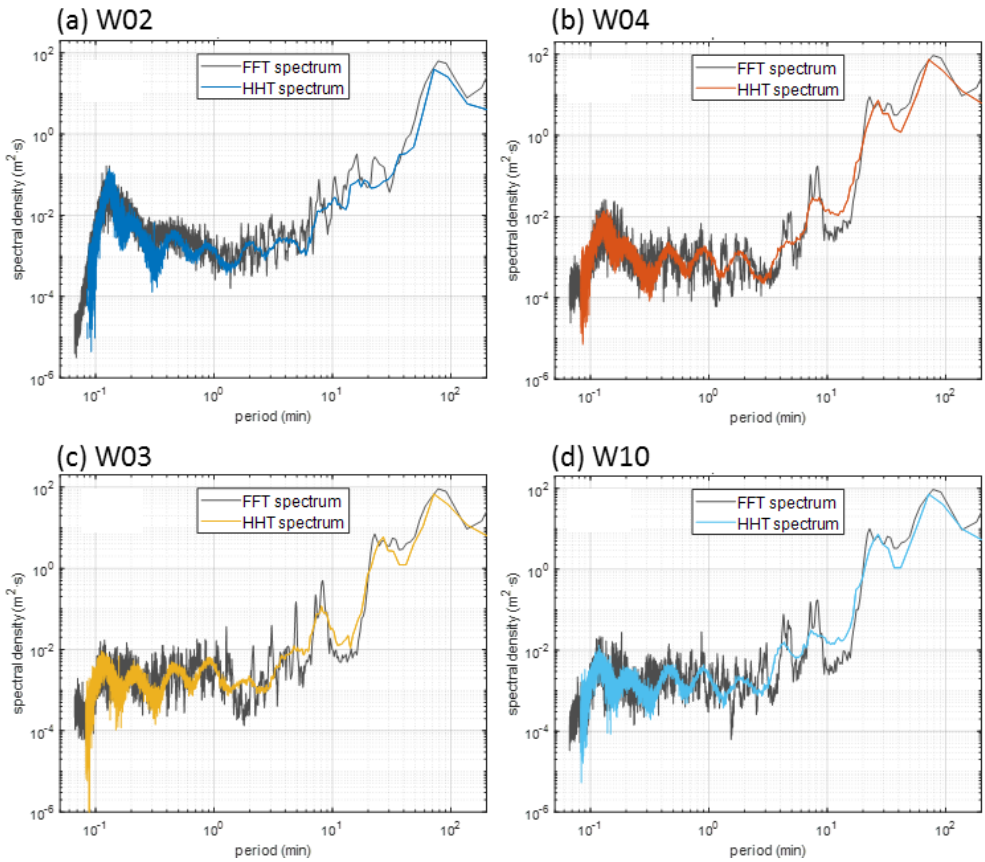


Figure 4.3 Comparison of marginal HHT spectrum and Fourier spectrum (black line) at the four wave stations on Mar.2, 2009. (a) at W02 (blue line for HHT spectrum). (b) at W04 (red line for HHT spectrum). (c) at W03 (yellow line for HHT spectrum). (d) at W10 (light blue line for HHT spectrum).

4.2 HHT Analysis of Sea Surface Elevation

Because the HHT analysis has an advantage of illustrating the spatio-temporal energy variation of a given signal, this methodology can be effectively used for disclosing the intrinsic mechanism of harbor agitation and the resulting downtime events occurred in a harbor. In this respect, the coincidence and relevance of the observed waves with the downtime events in Pohang New Port were investigated by using the HHT analysis.

In order to examine the spectral energy variation following the downtime occurrences, HHT analysis was conducted for the resampled SSE data of 2-seconds interval for the downtime events listed in Table 4.1. In the report of PRMAPO (2010), the causes of the downtime at Pohang New Port were analyzed based on the observed waves on site and wind data provided by the Korea Meteorological Agency (KMA). The report assigned the causes of downtime occurrence and continuation separately depending on different reasons such as swell, wind sea, and strong wind. As shown in the table, the downtime events mainly occurred in winter, and "swell" was the primary cause, especially for the continuation of the downtime events.

Table 4.1 Downtime events listed in PRMAPO (2010)

Downtime period	Quay No.	Cause of downtime	
		occurrence	continuation
Nov.09, 2008 20:00 ~ Nov.11, 2008 13:00	7, 8	WS	WS+Swell
Nov.28, 2008 16:00 ~ Nov.29, 2008 10:00	7, 8	Swell	Swell
Nov.30, 2008 16:00 ~ Dec.01, 2008 07:00	7	Swell	Swell
Dec.14, 2008 01:00 ~ Dec.15, 2008 07:00	7, 8	Swell	Swell
Dec.18, 2008 21:00 ~ Dec.20, 2008 07:00	7, 8	Swell	Swell
Dec.21, 2008 19:00 ~ Dec.23, 2008 07:00	7, 8	Swell	Swell
Dec.28, 2008 13:00 ~ Dec.29, 2008 07:00	7, 8	Swell	Swell
Dec.31, 2008 08:00 ~ Jan.04, 2009 18:00	7, 8	Swell	Swell
Jan.09, 2009 18:00 ~ Jan.15, 2009 12:00	7, 8	Swell	Swell
Jan.30, 2009 08:00 ~ Feb.02, 2009 07:00	7, 8	SW+WS	Swell
Feb.15, 2009 16:00 ~ Feb.16, 2009 07:00	7	Swell	Swell
Feb.18, 2009 04:00 ~ Feb.18, 2009 24:00	7, 8	Swell	Swell
Feb.23, 2009 21:00 ~ Feb.23, 2009 10:00	7, 8	WS	Swell
Feb.26, 2009 01:00 ~ Feb.26, 2009 07:00	7	Swell	Swell
Mar.02, 2009 08:00 ~ Mar.02, 2009 18:00	7	WS	WS
Mar.19, 2009 19:00 ~ Mar.21, 2009 07:00	7, 8	SW+WS	Swell
Mar.23, 2009 07:00 ~ Mar.24, 2009 10:00	7, 8	Swell	Swell
Apr.01, 2009 10:00 ~ Apr.02, 2009 10:00	7, 8	Swell	Swell
Apr.14, 2009 08:00 ~ Apr.15, 2009 07:00	7, 8	SW+WS	Swell
Apr.16, 2009 16:00 ~ Apr.17, 2009 07:00	7, 8	SW+WS	Swell
Apr.24, 2009 10:00 ~ Apr.24, 2009 22:00	7	Swell	Swell
Apr.26, 2009 10:00 ~ Apr.29, 2009 15:00	7, 8	Swell	Swell

*SW = Strong wind, WS = Wind sea

4.2.1 Temporal Variation of the HHT spectra

Increase of energy in GW band during downtime occurrence

Figure 4.4 shows the HHT spectra during Jan.31, 2009 ~ Feb.02, 2009 (the third downtime event in Table 4.1) at the four wave stations (W02, W04, W03, and W10). Each of the four diagrams in the figure comprised of two panels of HHT spectra corresponding to relatively shorter and longer period bands. In the lower four panels, high energies are observed in GW band (0.5~5 minutes) at all the stations.

Meanwhile, the time series of wave data during the downtime event is illustrated in Figure 4.5 to seek more synthetic understanding of the temporal variation of wave field during the downtime event. The first two diagrams in Figure 4.5 show the time series of SSE at W03 and W10 stations during Jan.30, 2009 ~ Feb.02, 2009. The third diagram of the figure shows the time series of H_s and H_{sl} at the two wave stations. Also shown in the diagram is the modeled H_s at C01, whose location is outside the port as shown in Figure 3.2. During the downtime occurrence, illustrated as green shading in the figure, the magnitude of SSE and wave heights are significantly amplified. The HHT spectral energy in IGW band (0.5~5 minutes) was not significantly increased as shown in Figure 4.4, but H_{sl} significantly grown in Figure 4.5 during the downtime event.

Such a discrepancy can be reconciled by investigating the daily marginal HHT spectra shown in Figure 4.6, which is obtained by taking temporal aggregation of the HHT spectra during the own day. In this figure, DT=1 denotes that the spectra correspond to the day of downtime occurrence, while DT=0 did not. The four shaded bands in the figure show the time periods of 0.05~0.5, 0.5~5, 20~35, and 70~95 minutes, respectively, which are associated with the typical GW/IGW period and several energy

peaks of the spectra. It is obvious that the energy increase in GW band is more prominent than that in IGW band. Although wave energies in both GW and IGW bands increased during the downtime, the former was comparatively more apparent. Hence, the temporal evolution of IGW energy was not so much noteworthy in the HHT spectra in Figure 4.4, while both H_s and H_{sl} increased in Figure 4.5. This type of information cannot be obtained by the conventional Fourier spectral analysis that is suitable for analysis of stationary data, which demonstrates the reason why HHT analysis should be used to recognize crucial factor affecting the downtime event in the present study.

Comparison of the marginal HHT spectra on January 30th (DT=0) and on February 2nd (DT=1) shows that GW energy at around 0.2 minutes (12 seconds) apparently increased with the downtime occurrence whereas energies at IGW and NOP band are not. Hence, they seemed to have no association with the downtime event, indicating GW would be the sole factor for generating the downtime event.

Increase of energy in both GW and IGW bands during downtime occurrence

Figure 4.7 shows the HHT spectra during Dec.30, 2008 ~ Jan.05, 2009 (the second downtime event in Table 4.1) at the four wave stations. In this case, temporal variation of the HHT spectra at the W03 station in the third diagram of the figure exhibits significantly enhanced energies in both GW and IGW bands during the downtime event. Spectral energies at around 1 and 4 minutes within IGW band apparently rise in the midst of downtime duration, which is not displayed in the HHT spectra during Jan.30, 2009 ~ Feb.02, 2009 shown in Figure 4.4.

Figure 4.8 shows the time series of SSE, H_s, and H_{sl}, where the almost similar variational trend of H_s and H_{sl} was observed as in the previous downtime event in

Figure 4.5. However, the daily marginal HHT spectra shown in Figure 4.9 illustrate clear difference from those in Figure 4.6. In this case, the spectral energy levels in both GW and IGW bands are enhanced during the downtime, and their magnitudes are almost comparable each other. Within the IGW band, the double energy peaks around 1 and 4 minutes are clearly displayed, which corresponds to the development of high energy zones at around those periods during the downtime as shown in Figure 4.7. Such amplification of IGW energy appeared only at the two stations of W03 and W10. According to van der Molen (2006), the NOP of mooring lines of the vessels moored at the slips is similar as the IGW period. Indeed, occasional long-period vessel oscillations or seiching has been consistently reported by the workers at the slips. In this context, spectral energy of IGW can increase to the level significantly affecting the agitation of moored vessels, although GW component is still the more dominant factor for involving downtime events.

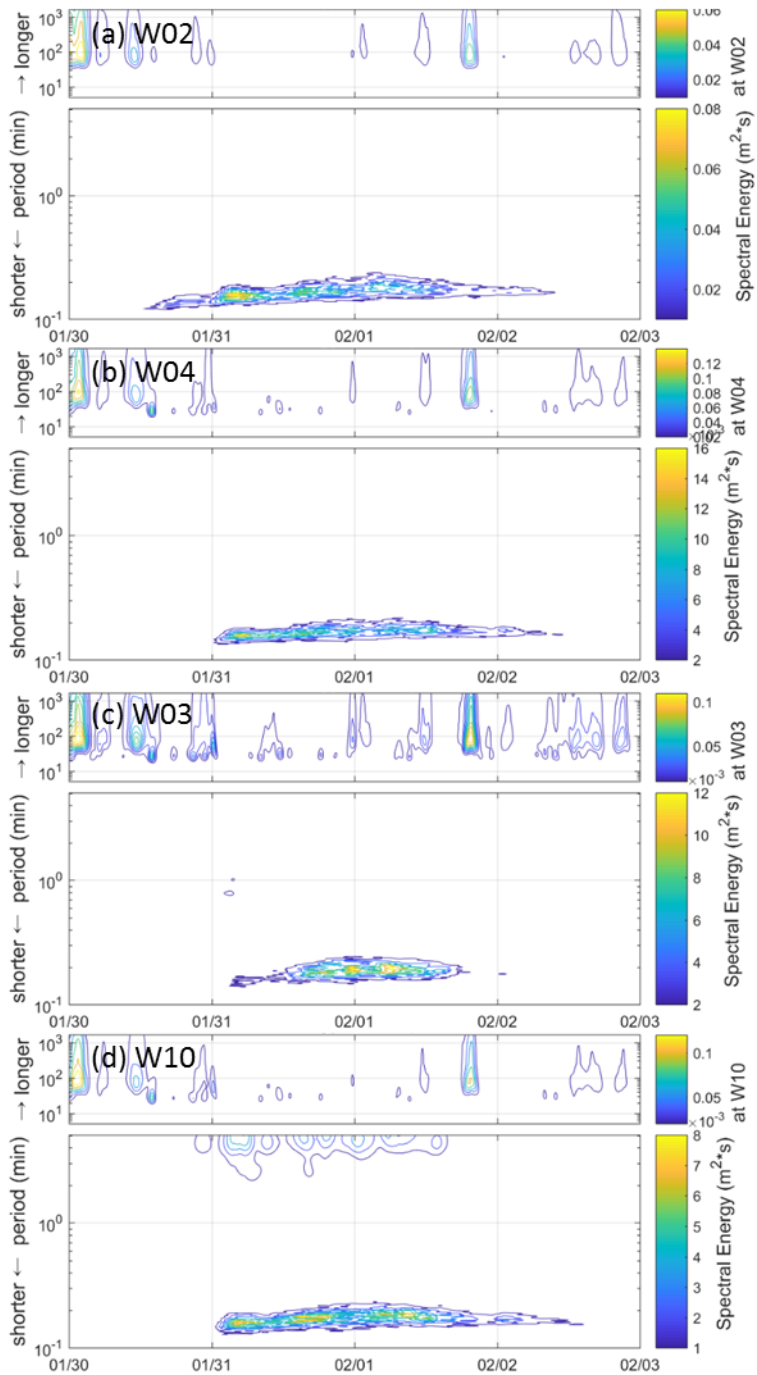


Figure 4.4 The time variation of HHT spectra at the four wave stations from Jan.30, 2009 to Feb.02, 2009. The period band of the upper panel on each plot is related to the NOPs of the port. The period of the lower panel include the period band of GW and IGW.

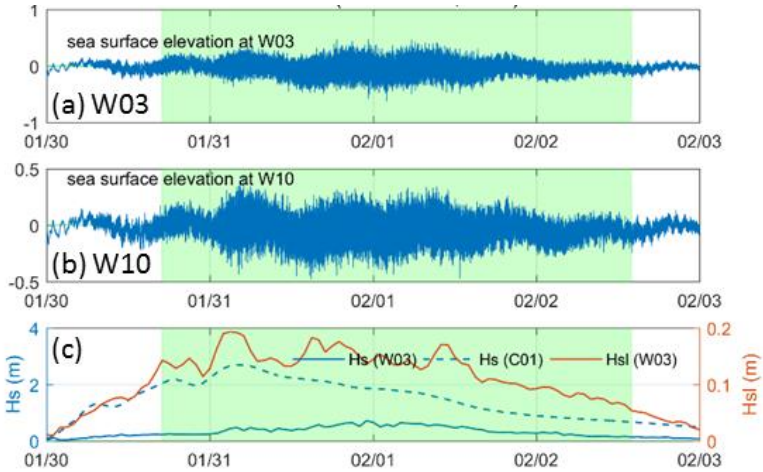


Figure 4.5 (a), (b) The SSE data at W03 and W10 (Jan.30, 2009 ~ Feb.02, 2009). (c) the significant wave heights of GW and IGW at W03 with that of GW at C01 for the same duration.

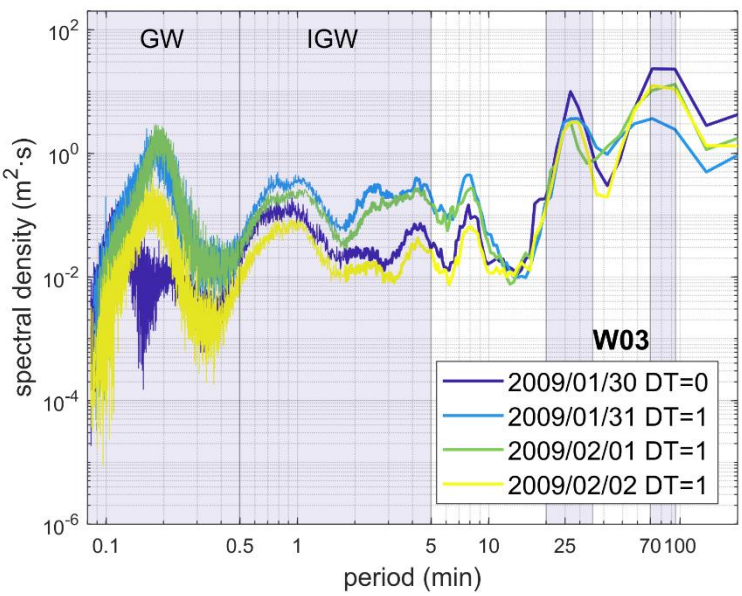


Figure 4.6 Comparing marginal HHT spectra at W03 for each one-day during Jan.30, 2009 ~ Feb.02, 2009. "DT=1" denotes that downtime occurred at that day, but "DT=0" means that there was no downtime occurrence at that day.

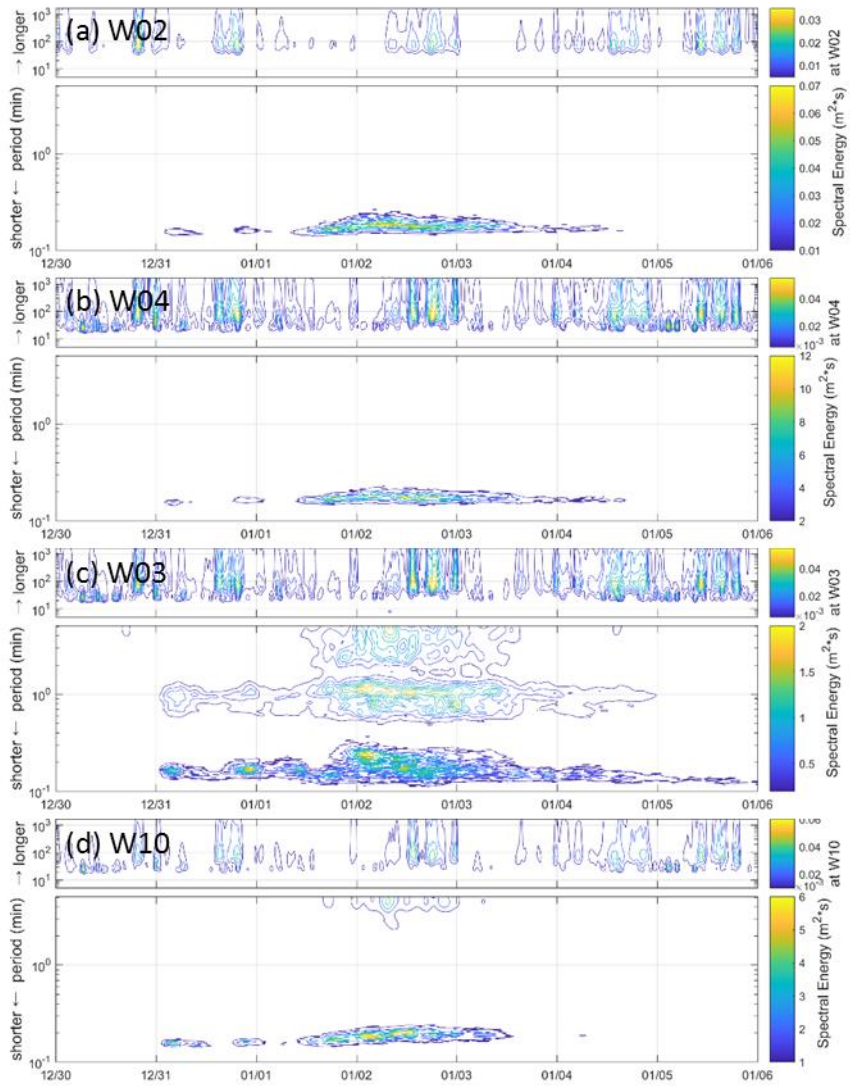


Figure 4.7 The time variation of HHT spectra at the four wave stations from Dec.30, 2008 to Jan.05, 2009. The period band of the upper panel on each plot is related to the NOPs of the port. The period of the lower panel include the period band of GW and IGW.

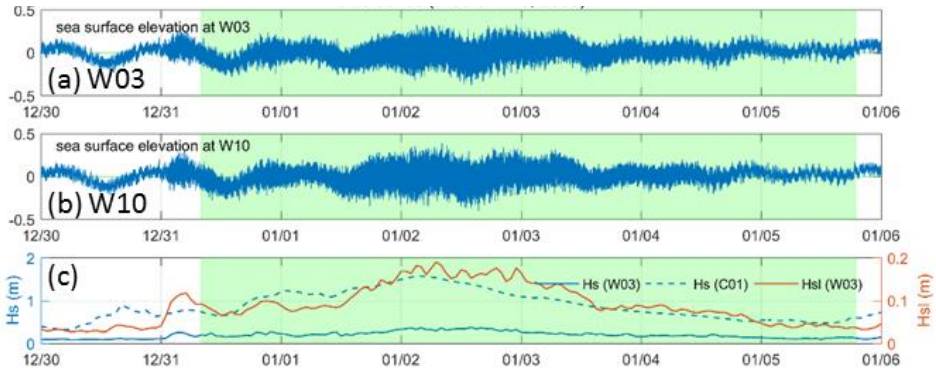


Figure 4.8 (a), (b) The SSE data at W03 and W10 (Dec.30, 2008 ~ Jan.05, 2009). (c) the significant wave heights of GW and IGW at W03 with that of GW at C01 for the same duration.

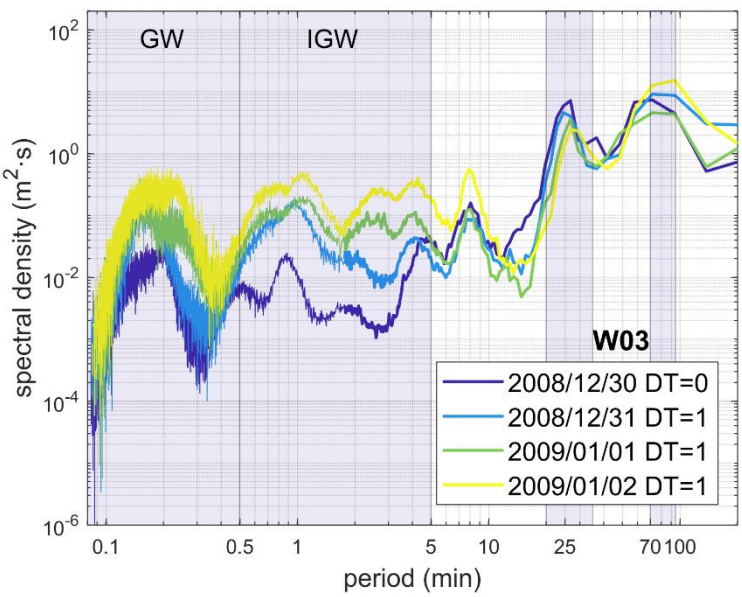


Figure 4.9 Comparing marginal HHT spectra at W03 for each one-day during Dec.30, 2008 ~ Jan.02, 2009. "DT=0" denotes that downtime occurred at that day, but "DT=1" means that there was no downtime occurrence at that day.

4.2.2 Comparison of HHT Spectra at Multiple Stations

The Marginal HHT spectra during one day at the four stations are plotted in Figures 4.10 to 4.13, where DT=1 denotes that the spectra correspond to the day of downtime occurrence, while DT=0 did not. The four shaded bands in the figure show the time periods of 0.05~0.5, 0.5~5, 20~35, and 70~95 minutes, respectively, which are associated with the typical GW/IGW period and several energy peaks of the spectra.

In those figures, regardless of the downtime occurrence and the location of stations, a remarkably high energy peak is observed at the period in between 70~95 minutes, which is related to the natural oscillation period (NOP) of Youngil bay. The another high peak appears at around 25 minutes except for W02 station, which indicates the close association of the peak with NOP of Pohang New Port, because W02 station is the only station outside the port. Moreover, the period of NOPs inside the slips is in the range of IGW. Another noteworthy thing about the figures is that energies of both NOPs change insignificantly irrespective of the downtime occurrence. Their spectral energies are almost the same for the marginal HHT spectrum with DT=0 and DT=1. This fact indicates that both NOPs are not relevant to the downtime events.

In contrast, the spectral energies within GW (less than 30 seconds) and IGW (period between 0.5 to 3 minutes) bands significantly increased with the downtime occurrence because of the marginal HHT spectra with DT=1 show clearly higher energy levels than those with DT=0. More specifically, the marginal HHT spectrum at W02 station shows apparently higher energy in GW band than those at the other stations. This aspect indicates that wave energy in GW band is reduced after waves propagate into the port. In contrast, the spectral energies of IGW at W03 and W10 stations are comparatively

higher rather than those at the other two stations. It is a probable aspect because the bound IGW is unlocked and energies are transferred from GW to IGW band inside the port. For this reason, the spectral energy levels at W03 and W10 that are located deep inside the basin (see Figure 3.3) can be increased as an influence from the IGW-induced agitations.

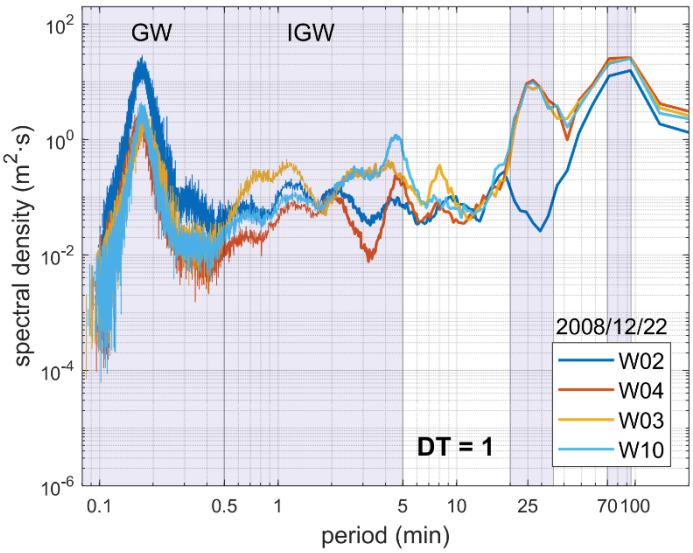


Figure 4.10 Comparison of marginal HHT spectra at multiple stations W02~W10 on Dec.22, 2008. "DT=1" denotes downtime occurrence on the day.

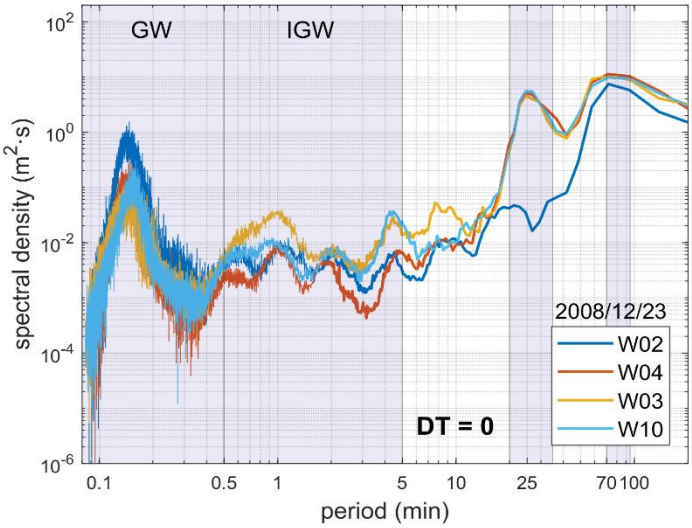


Figure 4.11 Comparison of marginal HHT spectra at multiple stations W02~W10 on Dec.22, 2008. "DT=0" denotes no downtime occurrence on the day.

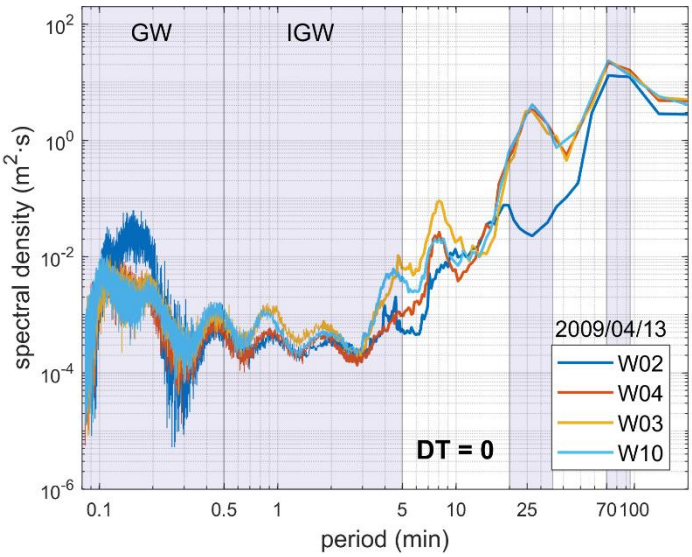


Figure 4.12 Comparison of marginal HHT spectra at multiple stations W02~W10 on Apr.13, 2009. "DT=0" denotes no downtime occurrence on the day.

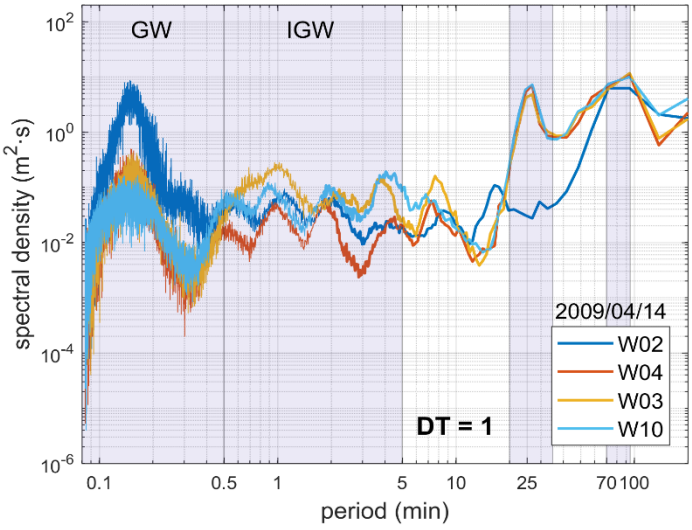


Figure 4.13 Comparison of marginal HHT spectra at multiple stations W02~W10 on Apr.14, 2009. "DT=1" denotes downtime occurrence on the day.

4.3 Quality Control of Downtime Data by Using HHT Analysis

The HHT analysis could further be used for examining the adequacy of the downtime data. The downtime data was made from the downtime records that had been written manually by the workers at the quays of the port. Hence they may involve some human errors, which will have an influence on the analysis results in this study. In this context, the adequacy of the downtime data was carefully inspected, and the HHT analysis played a key role in the inspection and subsequent modification of some erroneous data included in the original downtime data. Such quality control of the downtime data contributed to improving the performance of the classification analysis that will be explained in Chapter 6.

4.3.1 Necessity of Examining Quality of Downtime Data

The downtime data are produced from the collection of the downtime charts written daily at the quays of Pohang New Port. These charts contain all the details, including quay locations, name and weight of vessels, and hourly records when a downtime started and ended. The downtime records were manually written by different workers while they were working at a certain quay. For this reason, the records may contain human mistakes or systematical errors that were not well recognized at the time of recording. The common mistakes or errors that can be included in the downtime records are as follows.

Irreverence of downtime records at neighboring quays

In the case wherein one quay has many downtime records while another have relatively few downtime records, it may be thought that downtime events frequently occur in the former quay and harbor agitation scarcely occurs at the latter quay. However, a quay would have few downtime if the wave condition around the quay was already so severe that one did not even try to load or unload cargoes at the quay owing to the associated operational costs. Therefore, instead of judging the downtime conditions based only on the downtime data at a single quay, it is desirable to check the downtime situation by considering overall wave conditions and the downtime records obtained at other quays.

Wrong recording of stopping and resuming time of operation

The start and end time of a recorded downtime may not be accurate as they depend on the recorder's handwriting. It is possible that the stopping time of operation was recorded delayed than the actual time. Also, it is not fair to say that the time of resuming the operation was recorded right after the port condition had been sufficiently stabilized. Even if the circumstance became favorable, the operation could not be resumed because it was difficult to restart the operation due to other reasons. For instance, the downtime might be recorded as continued up until the dawn and was relieved in the morning even though the operation was possible during the midnight.

4.3.2 Modification of Downtime Data

Considering the necessity of quality control of the downtime data, adequacy of the data was examined based on the HHT analysis results. Quality control of the data was carried out for the downtime data obtained at W03, W04, and W10 stations that are located on Pier 7 and 8 (see Figure 3.3), where downtime events most frequently occurred within the port.

As explained in Section 4.2, energy increase of GW or IGW in a HHT spectrum is not always correspondent to the occurrence of downtime. In some cases, for example, the downtime records continued until the morning even the wave condition according to HHT spectrum became mild enough to resume operation during the midnight. Also, in some other cases, the starting time of downtime seemed to be delayed far after a significant increase of the wave energy in GW or IGW bands in the HHT spectrum. Due to these types of discrepancies between the HHT spectrum and the downtime records, all the downtime data were carefully compared with the corresponding HHT spectrum and then subjected to one of the three following treatment of the data.

1. Confirmation: the original index of downtime (DT=1 or 0) was confirmed if there is a good agreement between the HHT spectrum and the downtime records.

2. Modification: the original index of downtime (DT=1 or 0) was converted to the opposite index if it is obvious that the downtime records are mistakenly written in view of the spectral energy levels of GW and IGW in the HHT spectrum at the time instant.

3. Elimination: the original index of downtime (DT=1 or 0) was made null if there is notable disagreement between the HHT spectrum and the downtime records, but it is not firmly clear that the downtime records are mistakenly written or not. The

corresponding downtime data was no more used as input data for the classification analysis that is described in Chapter 6.

More specifically, to quantitatively identify and deal with some possible erroneous data included in the original downtime records, statistical quantities from the wave data of H_s and H_{sl} are also used together with the HHT spectrum analysis. Figure 4.14 shows boxplots and histograms of H_s and H_{sl} at W03 station for two different groups of data corresponding to DT=0 and DT=1, respectively. Because the values of H_s and H_{sl} with DT=1 were much higher than those with DT=0, it is very natural that the histograms of H_s and H_{sl} show completely different distributions depending on DT=1 or DT=0.

According to inter-quartile rule for outliers, the data outside the value of Q3 +1.5x(Q3-Q1) can be regarded as outliers, where Q1 and Q3 denote the first and third quartile, respectively. In this study, this rule was applied as a threshold for identifying possible erroneous downtime data indexed with DT=0. The threshold values for H_s and H_{sl} at W03 were calculated to be H_s=0.213 m and H_{sl}=0.077 m, which are illustrated as pink lines in the box plots and histograms in Figure 4.14. The threshold values at the other two stations were obtained in the same way.; the calculated values are H_s=1.130 m and H_{sl}=0.039 m at W02 whereas H_s=0.258 m and H_{sl}=0.042 m at W04. Then a new provisional downtime time history was made having the index value of either DTp=0 or DTp=1, depending on the values of H_s and H_{sl} at a particular time instance are lower or higher than the thresholds.

As shown in Figure 4.15, the original downtime time histories and the new provisional downtime time histories were slightly different. For the time sections of having the same index value (DT=0 & DTp=0 or DT=1 & DTp=1), the original downtime data were confirmed. If they show different index value (DT=0 & DTp=1 or DT=1 & DTp=0), then spectral energy levels in the HHT spectrum during those time sections were examined to determine whether the data are modified or eliminated. In Figure 4.15, for example, the original downtime records of DT=0 during 00:30 to 04:00 on March 20th were adjusted to DT=1 because it is obvious that spectral energies of GW and IGW are significantly high at those times. Meanwhile, the original downtime records of DT=1 during 03:00 to 05:30 on March 21st were eliminated because it is more or less ambiguous if spectral energies of GW and IGW are apparently insignificant or not. A similar procedure was also applied to the following downtime event over March 23th to 24th.

In the time series of H_s and H_{sl} in Figure 4.15, the green shaded area represents the time spans corresponding to the original downtime records, whereas the vertical magenta lines denote the time instances of the revised downtime records including the confirmed and modified downtime data. The revised downtime data after experiencing the quality control contributed to enhancing the performance of downtime forecast by the classification method, which will be explained in Chapter 6.

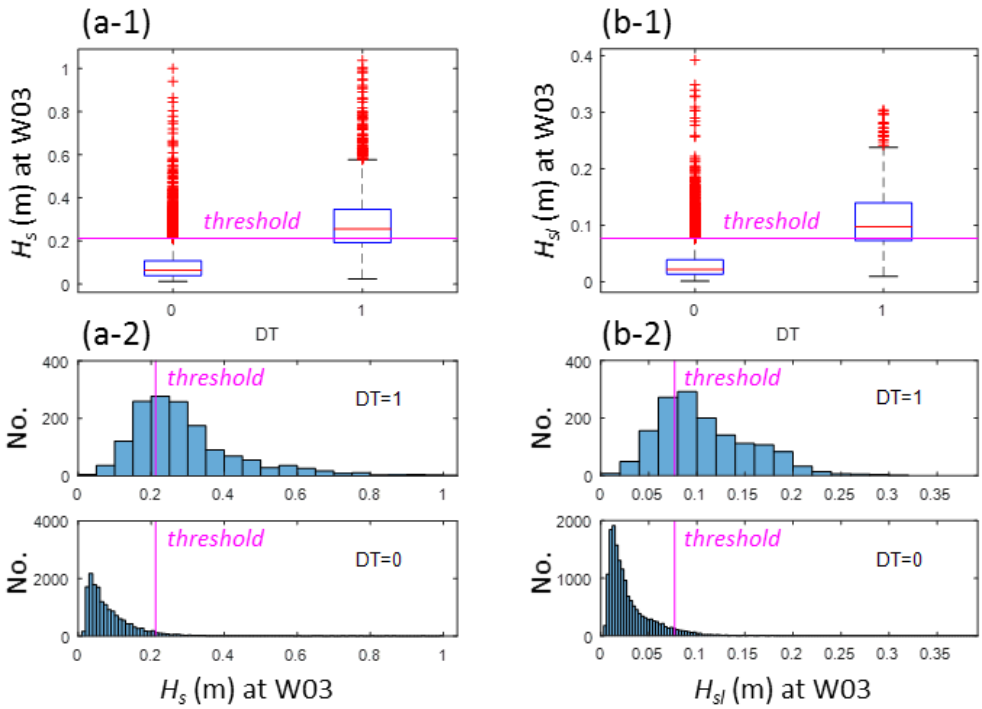


Figure 4.14 (a-1) Boxplot and (a-2) histograms of significant wave heights of GW, H_s at W03 in accordance with the data with DT=0 and DT=1. The threshold (H_s = 0.213 m) indicates the outlier of the data H_s that is coincide with DT=0. (b-1) Boxplot and (b-2) histograms of significant wave heights of IGW, H_{sl} at W03 in accordance with the data with DT=0 and DT=1. The threshold (H_s = 0.077 m) indicates the outlier of the data H_{sl} that is coincide with DT=0.

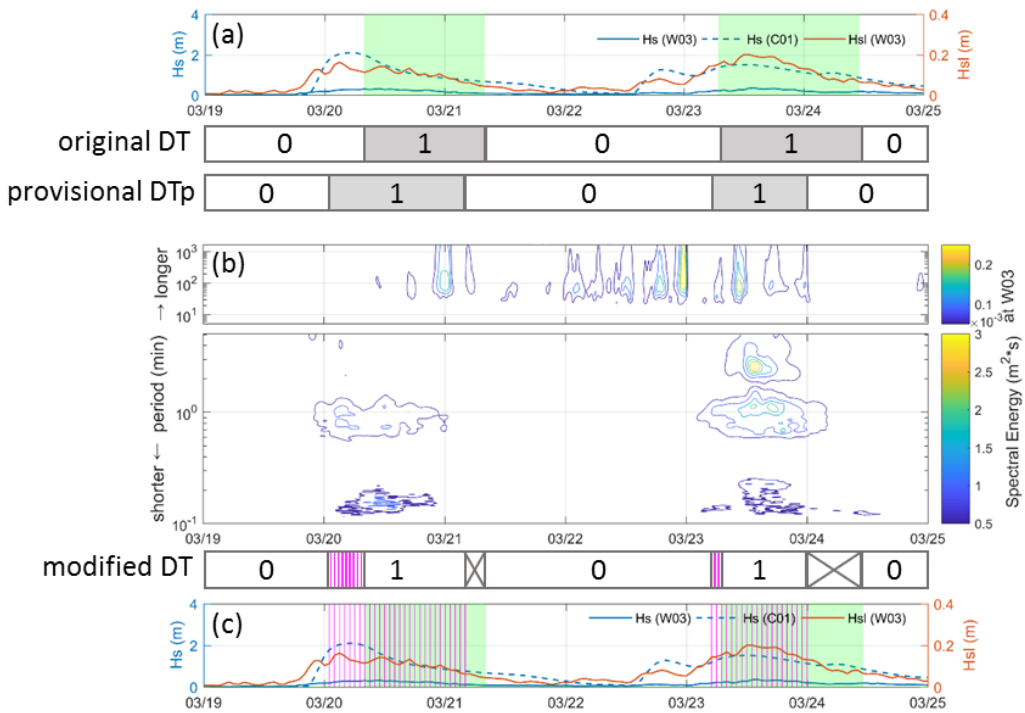


Figure 4.15 An example of quality control of the original downtime data during Apr.13, 2009 ~ Apr.17, 2009. (a) the significant wave heights of GW and IGW at W03 with the original downtime sequence. (b) the time variation of HHT spectra at W03. (c) (a) with the modified downtime sequence. The green shaded area represents the time spans corresponding to the original downtime records, whereas the vertical magenta lines denote the time instances of the revised downtime records including the confirmed and modified downtime data.

CHAPTER 5.

WAVE PREDICTION WITH NEURAL NETWORKS

Prediction of waves inside and outside harbor as well as inside the slip, which is used for classification as predictor inputs, is dealt with in this chapter. The prediction time span for wave forecasting must be sufficiently long to inform expected downtime for harbor operation, that is, more than 24–72 hours. Accordingly, for enough prediction time span, the wind field data provided by Japan Meteorological Agency (JMA) that can be forecasted continuously is used firstly. Subsequently, the wave data numerically calculated from the wind data are used as input variables for the target observed multi-station data in Neural Networks (NN). The best NN model selection strategy is applied, and ensemble NNs propose more generalized prediction results.

First, the validity of the wave forecasts inside the port, through the wave observation data outside the port, was reviewed by using the conventional method, and the performances of the NN model were compared with the nonlinear regression model. Second, this study performed model selection to find an improved and more optimized NN model. Lastly, the study applied various ensemble methods to gain more generalized prediction results from the NN.

5.1 Conventional Approach for Wave Prediction

Prior to the comparison of the results of the prediction models, the empirical formulas and regression analysis were investigated. This investigation may serve the physical basis of the predictive model. There have been several studies on the estimation of IGW based on the correlation with GW (*e.g.* Nelson *et al.*, 1988; Bowers, 1992; McComb, 2005; Lopez and Iglesias, 2013; Thiebaut *et al.*, 2013). Most of them suppose a high dependency of the concurrent IGW on the offshore or inshore GW and suggest empirical formulas for predicting IGW.

McComb *et al.* (2005),

$$H_{sl} = K\left(H_s^{1.1}\, T_p^{0.5}\right) + C \tag{5.1}$$

Bowers (1992),

$$H_{sl} = K\left(H_s^{1.11}\, T_p^{1.25}\right)/h^{0.25} \tag{5.2}$$

Nelson *et al.* (1988),

$$H_{s_l} = \begin{cases} 0.013 & \text{if } H_s T_p < 18 \text{ ms} \\ 8.05 \times 10^{-5}\left(H_s\, T_p\right)^{1.77} & \text{if } H_s T_p \geq 18 \text{ ms} \end{cases} \tag{5.3}$$

where K and C are the slope and the intercept. H_s, H_{sl} are significant wave height of gravity and infragravity waves. T_p and h are peak wave period and water depth.

In order to apply such a method to Pohang New Port, the relationship between GW and IGW among the four measurement stations is investigated in this study. The analysis of detecting correlation between GW and IGW was described in chapter 3.3.2. The

scatter plots of H_{sl} and H_s among observed data reveal quadratic or close-to-linear proportionality, but those among calculated and observed data show a much more ambiguous relationship (in Figure 3.10 and Figure 3.11).

All the conventional formulas were developed based on multivariate nonlinear regression models, for calculating the infragravity wave height. Among them, the model showing the best agreement with the measured infragravity wave height is formulated as follows.

$$H_{sl} = a\, H_s^b\ T_z^c + d \tag{5.4}$$

where the four coefficients a, b, c, and d are the regression coefficients. The scatter plots of observed and the predicted H_{sl} at W03 and W04 are shown in Figure 5.1.

Some previous studies have supposed a proportional relationship between the infragravity wave heights in a nearshore region and gravity wave heights in the offshore area (Tucker, 1950). However, the relationship between the two quantities changes very rapidly during the propagation of waves from offshore to the shore through the continental shelf, surf zone, and swash zone. Such a complicated change is caused by variation of topography and complex wave processes as well. In this context, it is necessary to understand how much the energy is exchanged between the GW and IGW and in which process the energy is lost the most. The above regression analysis reveals that some energy would be transferred from gravity wave band to infragravity wave band during the propagation of waves into the harbor.

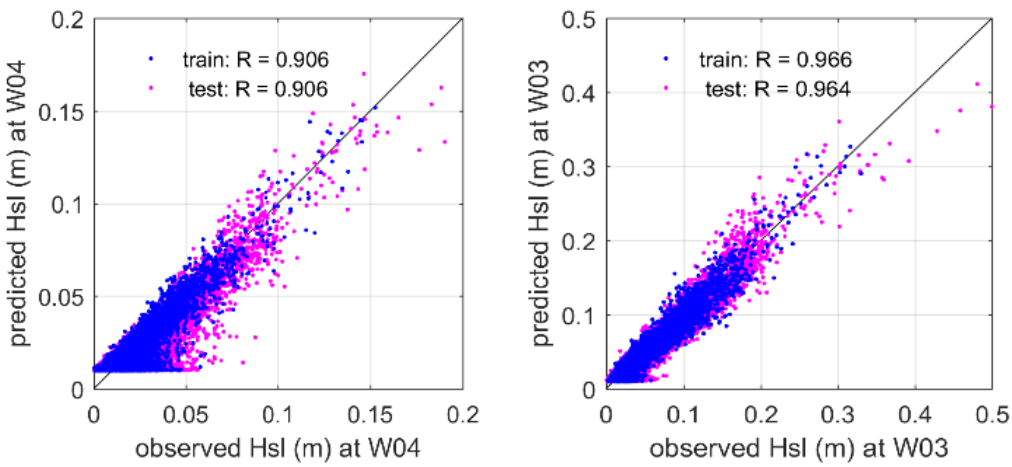


Figure 5.1 Scatter plots of predicted significant wave height of IGW, H_{sl} inside the harbor (at W03) by using significant wave height of GW, H_s outside the harbor (at W04).

5.2 Neural Network Prediction

In NN prediction models for waves observed at multiple-stations, target variables of wave characteristics are the significant wave heights of GW (H_s), IGW (H_{sl}), and mean wave period (T_z) at the three stations of W02, W03, and W04. Input variables from the calculated data are the significant wave height, peak period, wave direction, wind direction and wind speed. NN and regression analysis were adopted as prediction models.

Input and target data can be divided into train and data set regarding time. When *X* and *Y* are the input and target data sets, the train and test can be described as follows.

train set

$$X_{tr} = [x_1, x_2, \cdots, x_{N_{tr}}] \tag{5.5}$$

$$Y_{tr} = [y_1, y_2, \cdots, y_{N_{tr}}] \tag{5.6}$$

test set

$$X_{ts} = [x_{N_{tr}+1}, x_{N_{tr}+2}, \cdots, x_{N_{tr}+N_{ts}}] \tag{5.7}$$

$$Y_{ts} = [y_{N_{tr}+1}, y_{N_{tr}+2}, \cdots, y_{N_{tr}+N_{ts}}] \tag{5.8}$$

where N_{tr} is the length of train data and N_{ts} is the length of test data. The actual duration of train and test set is depicted in Figure 3.7.

A conceptual diagram of the prediction input and target is depicted in Figure 5.2 with the description of variables. Two different prediction schemes were constructed as illustrated in Figure 5.3 depending on the number of predicting steps. Target variables are directly predicted from the calculated wave data by the one-step prediction. Meanwhile, two-step prediction uses the results of previous step prediction as inputs for predicting the target. Either the NN or the regression can be utilized for each step prediction.

In Figure 5.4, the NN model shows better performance than the regression model especially regarding as the prediction of relatively higher wave heights, which are mostly under-predicted by the regression model. When the results from the one- or two-step prediction models are compared in Figure 5.5, the former shows better agreement with the observation in general. Figure 5.6 exhibits the time series of H_s at W04 that are predicted by the one-step NN model .

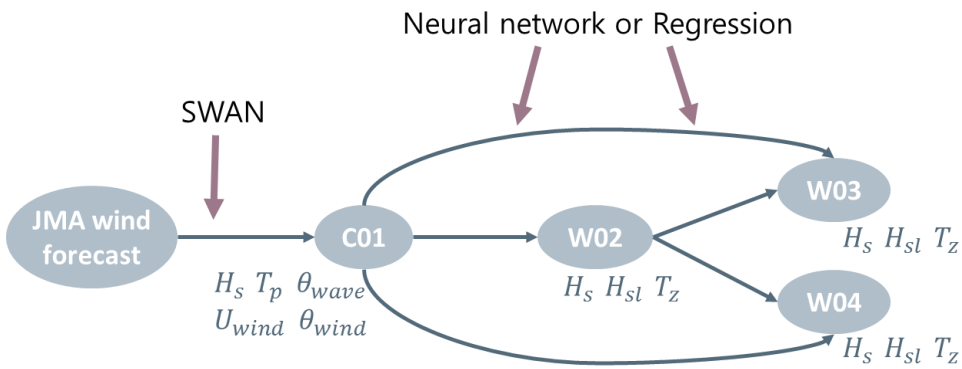


Figure 5.2 Conceptual diagram of wave prediction models

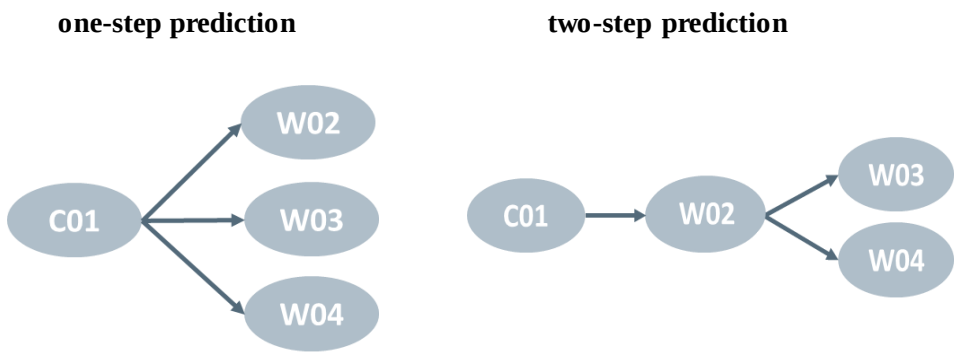


Figure 5.3 One- and two-step prediction scheme

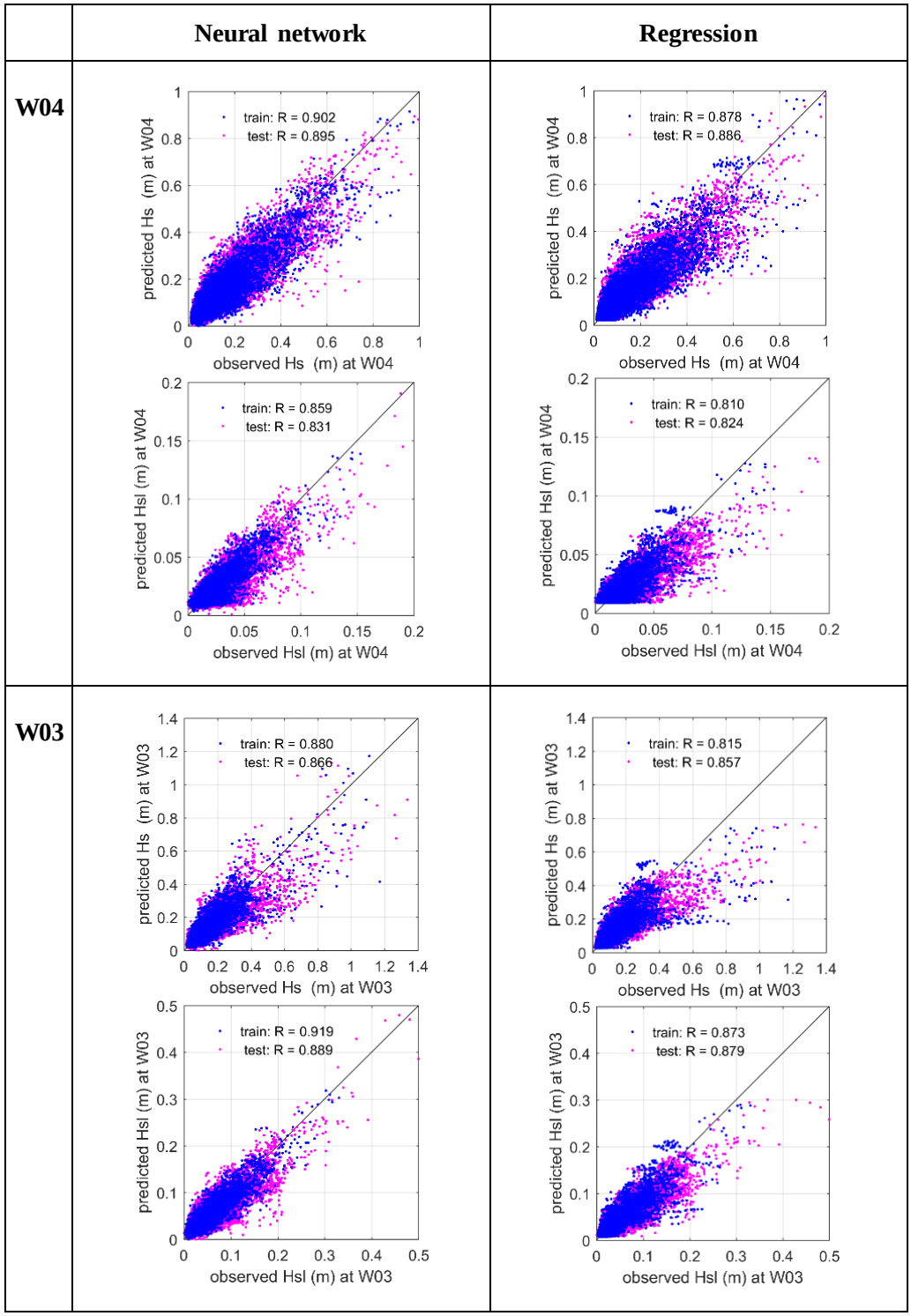


Figure 5.4 Performance comparison of one-step NN model and the nonlinear regression model

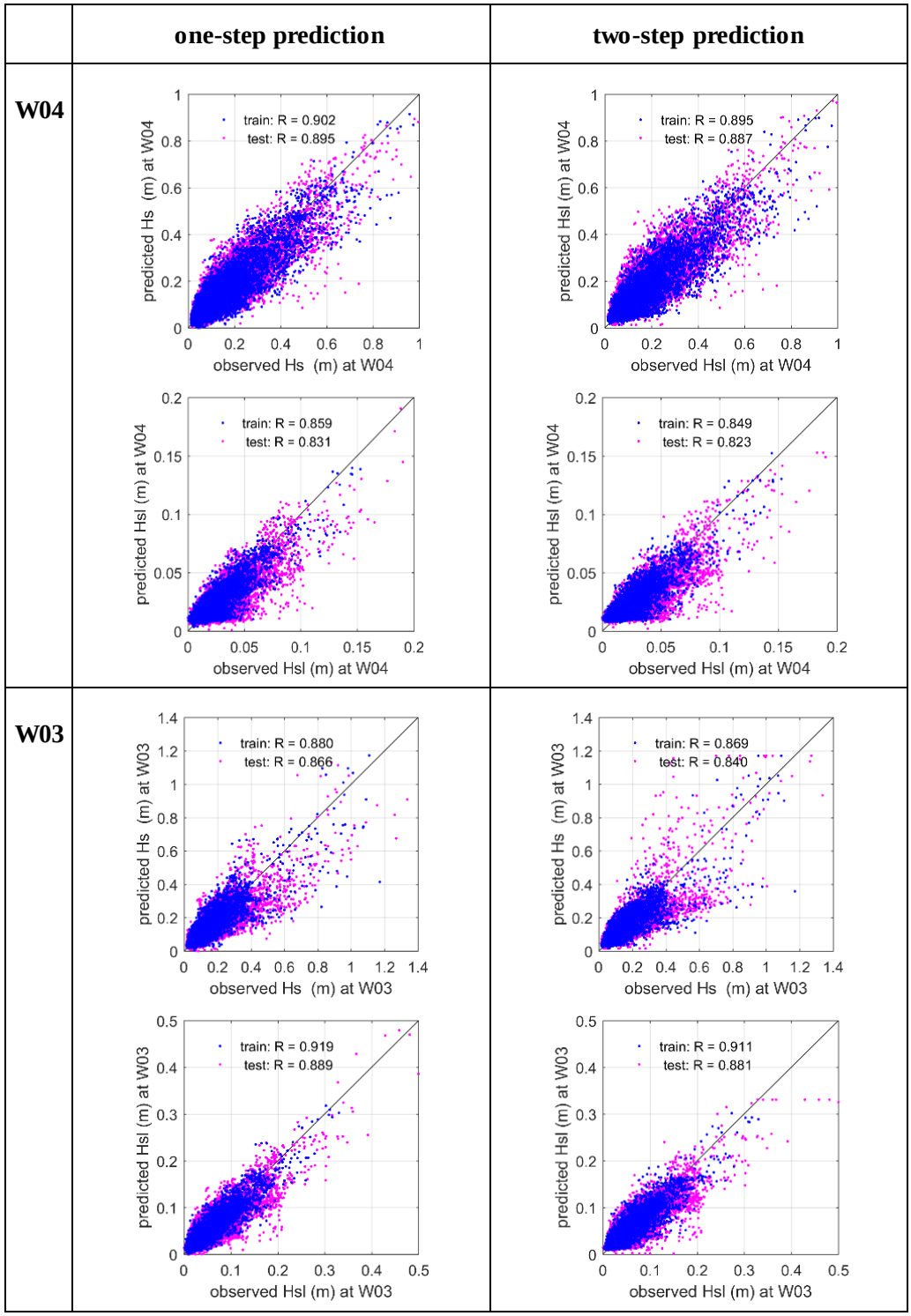


Figure 5.5 Performance comparison of one- and two-step NN prediction model

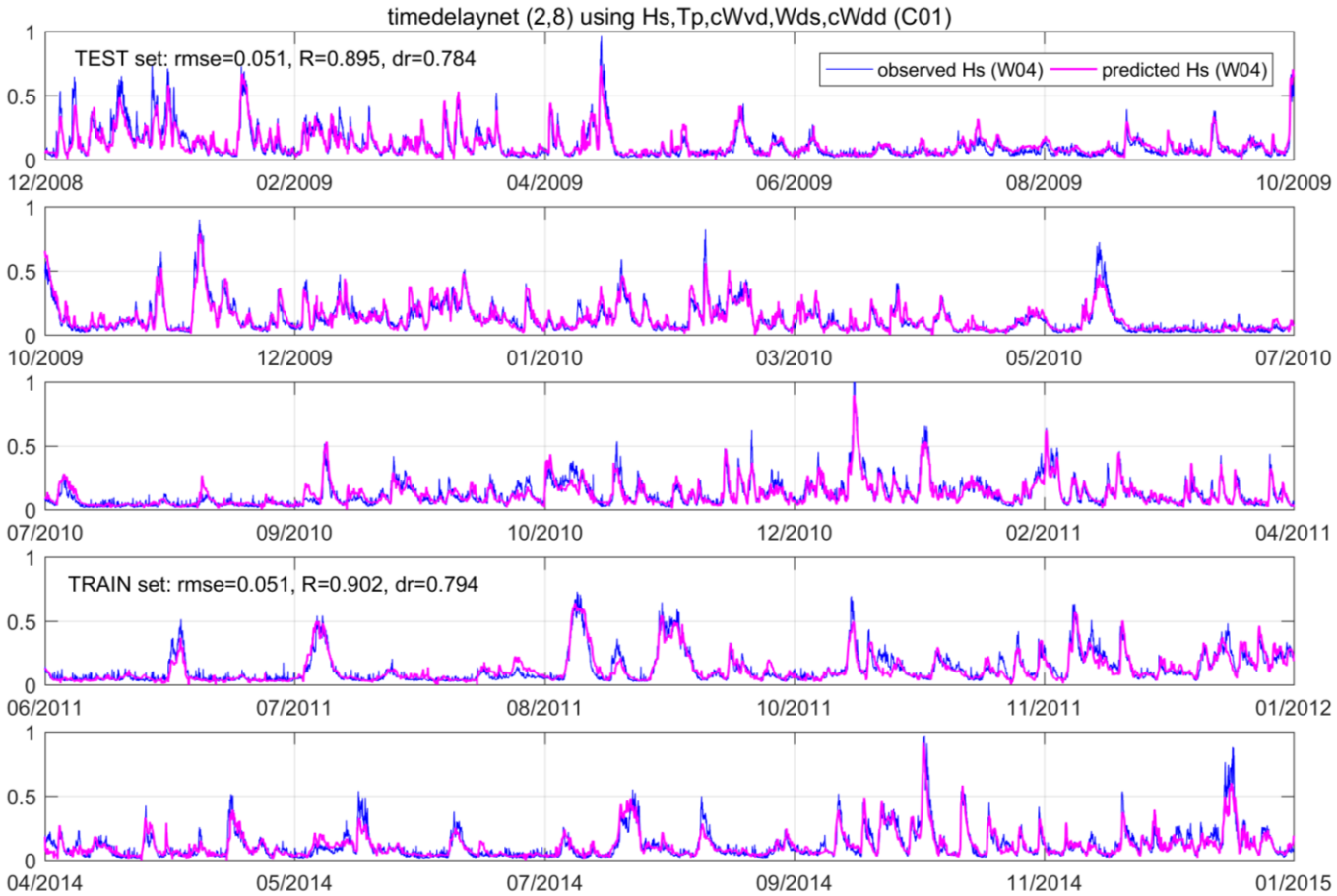


Figure 5.6 An exemplary time series showing performance of the one-step NN model

5.3 Model Selection and Ensemble of Neural Networks

5.3.1 Model Selection Strategy

As the NN in this study is for wave data modeling, the network architecture that fits the physical characteristics of the data is expected to deliver favorable and relevant results. This study compared criteria while changing the NN structures to gain the most optimized and generalized network. Various criteria to be applied to the forecasting model were examined, and it was observed that each had slightly different characteristics.

In order to test the validity of a model in model selection, the validation data set is introduced. In this study, the previous test set was divided into validation and test sets. As a result of model selection, the validation set was used for adjusting parameters of NN. The test set was only used for evaluation of new data.

Following the previous representation of test set in equation (5.7) and equation (5.8), the validation set (for $N_{vd} < N_{ts}$) is as follows

$$X_{vd} = [x_{N_{tr}+1}, x_{N_{tr}+2}, \cdots, x_{N_{tr}+N_{vd}}] \tag{5.9}$$

$$Y_{vd} = [y_{N_{tr}+1}, y_{N_{tr}+2}, \cdots, y_{N_{tr}+N_{vd}}] \tag{5.10}$$

In order to compare NN models and select the best model, the averaged performance of 100 trials of each model is estimated based on several criteria. Some criteria for model performance are introduced as follows.

Criterion for Model Performance

There are criteria using the error of model, such as SSE, MSE and so on. The basic, SSE is the sum of squares due to the error of the fit, and it indicates a fit that is more useful for prediction when a value closer to zero. MAE and MSE are quite helpful in evaluating the total absolute forecasting error, but they are susceptible to extreme errors as well as the change of scale and data transformations. RMSE is the root mean squared error or standard error, whose value closer to 0 indicates a fit that is more useful for prediction. RMSE has similar properties like MSE, but it is far less sensitive to the extreme errors and is more stable. NRMSE is the normalized RMSE; which facilitates the comparison between datasets or models with different scales. There are no consistent means of normalization in the literature, and the sum of squares using the mean value of the target is used in this study.

$$\mathrm{SSE} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.11}$$

$$\mathrm{MSE} = \frac{1}{N}\mathrm{SSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.12}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{5.13}$$

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{5.14}$$

where N is the number of data points of the time series that are considered, y_i and $\bar{y}$ denote the observed data and their mean value, and, $\hat{y}_i$ and $\bar{\hat{y}}$ are the modeled data and their mean value.

There are widely-used information-based criteria also, such as Akaike's information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC). They penalize large models that are likely to overfit. AIC is the most popular criteria for linear and nonlinear model identification. A common form of AIC is as follows

$$\mathrm{AIC} = \log\left(\hat{\sigma}^2_{\mathrm{MLE}}\right) + \frac{2m}{N} \tag{5.15}$$

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{\mathrm{SSE}}{N} = \mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.16}$$

where *m* is the number of parameters in the model. BIC weights greater penalty for the complexity of model than AIC. The use of BIC for model selection would be more helpful in a model with a relatively smaller number of parameters. One typical BIC is as follows

$$\mathrm{BIC} = \log\left(\hat{\sigma}^2_{\mathrm{MLE}}\right) + \frac{m\log(N)}{N} \tag{5.17}$$

All mentioned criteria measure different aspects of forecasting. Some researchers have suggested new indicators by summing some measures in appropriate manners (Egriloglu *et al*. 2008; Adhikari, 2015). However, this study identified the criteria from each of the train set and validation set and determined the best case in order to examine various aspects of the criteria. By comparing criteria of the models with different NN architectures, it was possible to determine whether the model of a certain NN architecture is generalized or over-fitted.

Setting Various Models

In the input layer, a total of 10 input neurons are selected with five input variables and two time delays. The output layer selects one target variable for each case. Compared to the layers themselves, the number of hidden layers and neurons can be set more freely. In general, the number of hidden layers, which controls the depth of the model, could sometimes make nonlinear data modeling more relevant. In this study, the number of hidden layers was set as one or two and that of hidden neuron was between two and ten.

In addition, the transfer function and training algorithm can be applied differently. The default transfer function is the tangent hyperbolic sigmoid function in the hidden layers whereas the linear function in the output layer. Meanwhile, two different training algorithm were used with Levenberg-Marquardt training backpropagation (termed as trainlm) and Bayesian regularization training backpropagation (termed as trainbr).

The procedure of model selection of the NN for target H_{sl} at W04 using input variables at C01 is described in Table 5.1. It compared the averaged performance of the networks after 100 iterations of each model. The average was calculated by excluding networks having outliers among the 100 iterations.

As shown in the table, the model performance for the train set is improved as the structure gets more complicated, or the number of parameters increases. On the other hand, in the test set, the model performance initially improves but later decreases even though the number of parameters increased. This is the moment where the model starts overfitting beyond the point of being generalized. In addition, it was possible to recognize the difference in performance among different algorithms and the number of

hidden layers. Based on this, a desirable NN architecture can be selected for each case from the results shown in Table 5.1. Figures 5.7 and 5.8 show the model selection process out of the candidate models.

Table 5.1 Model selection process for H_s (W03)

model	train NRMSE	test NRMSE	valid. NRMSE	train BIC	test BIC	valid. BIC	train R	test R	valid. R
Bayesian regularization training backpropagation algorithm									
10-2-1	0.528	0.525	0.537	-5.98	-6.01	-5.70	0.849	0.855	0.849
10-3-1	0.510	0.597	0.541	-6.04	-5.77	-5.67	0.860	0.803	0.844
10-4-1	0.494	0.566	0.540	-6.10	-5.85	-5.67	0.869	0.826	0.844
10-5-1	0.480	0.578	0.553	-6.15	-5.80	-5.61	0.878	0.820	0.835
10-6-1	0.469	0.589	0.571	-6.19	-5.76	-5.54	0.883	0.810	0.823
10-7-1	0.459	0.602	0.584	-6.22	-5.71	-5.48	0.888	0.801	0.813
10-[2-1]-1	0.528	0.528	0.529	-5.98	-6.00	-5.73	0.849	0.853	0.856
10-[3-1]-1	0.505	0.550	0.537	-6.06	-5.91	-5.68	0.863	0.840	0.847
10-[4-1]-1	0.482	0.566	0.545	-6.15	-5.85	-5.65	0.876	0.829	0.842
10-[5-1]-1	0.465	0.583	0.560	-6.21	-5.78	-5.58	0.885	0.817	0.831
10-[6-1]-1	0.454	0.590	0.568	-6.25	-5.75	-5.54	0.891	0.812	0.825
10-[7-1]-1	0.439	0.616	0.593	-6.31	-5.66	-5.45	0.899	0.794	0.809
Levenberg-Marquardt training backpropagation algorithm									
10-2-1	0.544	0.523	0.519	-5.92	-6.02	-5.77	0.839	0.859	0.866
10-3-1	0.523	0.526	0.518	-5.99	-6.00	-5.76	0.852	0.857	0.864
10-4-1	0.504	0.527	0.516	-6.06	-5.99	-5.76	0.863	0.856	0.862
10-5-1	0.491	0.539	0.524	-6.10	-5.94	-5.72	0.871	0.848	0.857
10-6-1	0.482	0.542	0.524	-6.13	-5.92	-5.70	0.876	0.846	0.856
10-7-1	0.476	0.554	0.535	-6.15	-5.87	-5.66	0.880	0.837	0.848
10-8-1	0.468	0.559	0.547	-6.18	-5.84	-5.60	0.884	0.834	0.839
10-[2-1]-1	0.528	0.528	0.529	-5.98	-6.00	-5.73	0.849	0.853	0.856
10-[3-1]-1	0.505	0.550	0.537	-6.06	-5.91	-5.68	0.863	0.840	0.847
10-[4-1]-1	0.482	0.566	0.545	-6.15	-5.85	-5.65	0.876	0.829	0.842
10-[5-1]-1	0.465	0.583	0.560	-6.21	-5.78	-5.58	0.885	0.817	0.831
10-[6-1]-1	0.454	0.590	0.568	-6.25	-5.75	-5.54	0.891	0.812	0.825
10-[7-1]-1	0.439	0.616	0.593	-6.31	-5.66	-5.45	0.899	0.794	0.809
10-[8-1]-1	0.431	0.658	0.613	-6.34	-5.52	-5.37	0.903	0.769	0.793

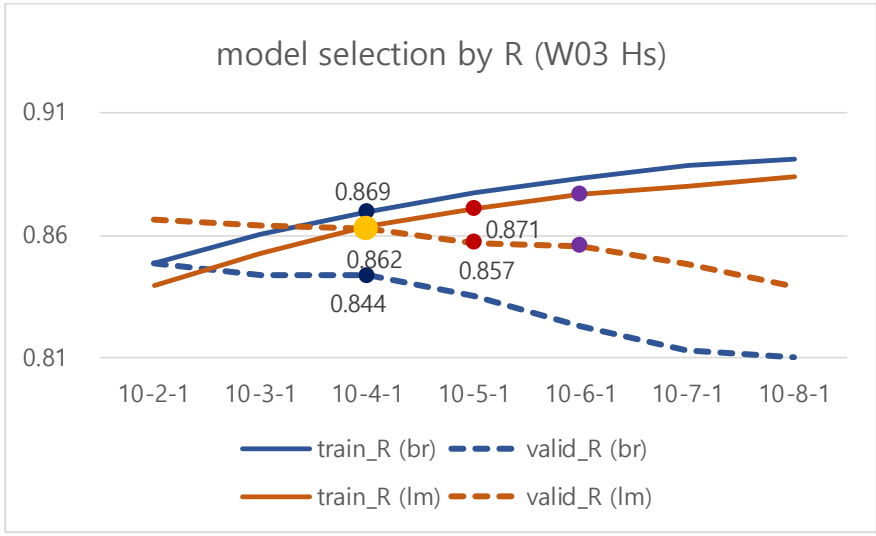


Figure 5.7 Comparison of R values from different models to select a best model for the train and validation sets

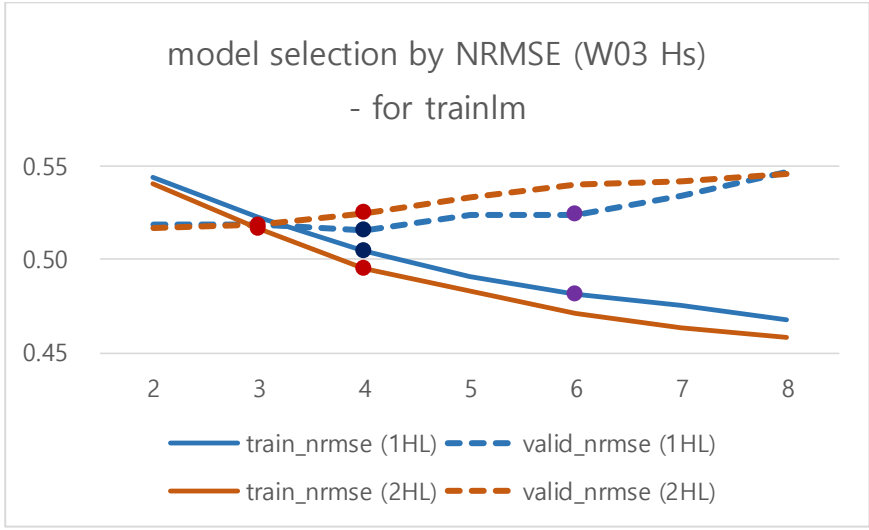


Figure 5.8 Comparison of NRMSE values from different models to select a best model for the train and validation sets

5.3.2 Ensemble Neural Networks

A neural network ensemble is another generalization scheme wherein a number of neural networks are trained under the best configuration through model selection. It is expected that such trials would lead to improved and more generalized results if the network forecasts are combined in appropriate ways. However, it is inevitable that the networks that were involved in the best model selection will have higher variance if the data is likely to be nonstationary and nonlinear. Therefore, when making an ensemble with several networks, it should be noted that not all the NN are likely to bring improvement. Zhou and Tang (2002) proved that an ensemble of many neural networks results in a better outcome than that of all the networks.

In this study, after training 100 networks of best selected NN model, some networks having a performance higher than a certain level defined by the previously mentioned criteria are found and ensembled properly for better prediction results. In other words, when 100 networks are obtained after 100 trials, intersections among these networks with a performance higher than a certain level of criteria in the validation sets can be determined and used for making an ensemble. In this study, NRMSE, BIC, and R are used as the criteria and quantiles and mean value of the criteria are used a certain level discriminating the model performance.

When trained nets using train set exist for *k*-th trial of 100 trials,

$$\hat{Y}_{tr}^{(k)} = net^{(k)}(X_{tr}) \tag{5.18}$$

we can get predicted $\hat{Y}$ for arbitrary input X using *k*-th network,

$$\hat{Y}^{(k)} = net^{(k)}(X) \tag{5.19}$$

That is, we can get a best individual forecast network as follows,

$$\hat{Y}^{(best)} = net^{(best)}(X) \tag{5.20}$$

The ensemble forecast can be described with the picked p networks, $net^{(j)} = net^{(1)}, \cdots, net^{p}$

$$\hat{Y}^{(ensemble)} = F\left(net^{(1)}(X), \cdots, net^{(p)}(X)\right) \tag{5.21}$$

Then, finally we can determine the ensemble operating function F using X_{tr}, Y_{tr} and $\hat{Y}_{tr}$

Ensemble operation

There are many ways to make ensembles with NNs. This research employed three different methods of the simple averaging method, the error-based method, and the linear combination method.

The Simple averaging method

This method has the virtues of impartiality and robustness. Hence many researchers have chosen this method consistently in broad fields (Clemen, 1989).

$$\hat{Y}^{(ensemble)} = \frac{1}{p}\sum_{j=1}^{p} net^{(j)}(X) \tag{5.22}$$

The error-based method

This method creates ensembles by adding weight on each network in inverse proportion to the error of each network (Armstrong, 2001).

$$\hat{Y}^{(ensemble)} = \sum_{j=1}^{p} ew_j \, net^{(j)}(X) \tag{5.23}$$

$$ew_j = e_j^{-1} / \sum_{j=1}^{p} e_j^{-1} \tag{5.24}$$

where e_i denotes the forecast error of the i-th network. The weight is a dimensionless coefficient. It is important to choose a specific error criteria for calculating the error-based weighs of ew_i.

The linear regression method

This method uses forecasts factors of each network as regressors of the ordinary least squares regression. Granger and Ramanathan (1984) argued that this method gives relatively less biased combined forecasts even if their factors are somewhat biased.

$$\hat{Y}^{(ensemble)} = \sum_{j=1}^{p} \left(a_j \, net^{(j)}(X) + b_j \right) \tag{5.25}$$

By using the three methods, the ensemble operation to the picked NN was conducted as described in Table 5.2. Q-0, Q-0.05, Q-0.1, and Q-0.5 represent that the standard to pick networks is described by quantile that corresponds to a probabilities of 0, 0.05, 0.1, and 0.5 respectively. MN, EW, LC indicate that the applied ensemble operator is the simple averaging method, the error-based method, and the linear regression method.

The method selected to create ensembles has a different effect on each case. Figure 5.9 shows comparison of the forecast improvement by using different ensemble operations. The combination using the linear regression has an advantage when the number of picked networks are small. There are other cases that do not show significant progress with ensemble operation. The ensemble results using the best selected model of all target data are shown in Table 5.3. Meanwhile, the most effective ensemble models for H_s and H_{sl} are shown in Figures 5.10 and 5.11, respectively.

Table 5.2 Notation of ensemble methods

Ensemble operation	cut-off values for sifting NNs				
	Q-0	Q-0.05	Q-0.1	Q-0.5	Mean
simple averaging method	MN1	MN2	MN3	MN4	MN5
error based method	EW1	EW2	EW3	EW4	EW5
linear combination method	LC1	LC2	LC3	LC4	LC5

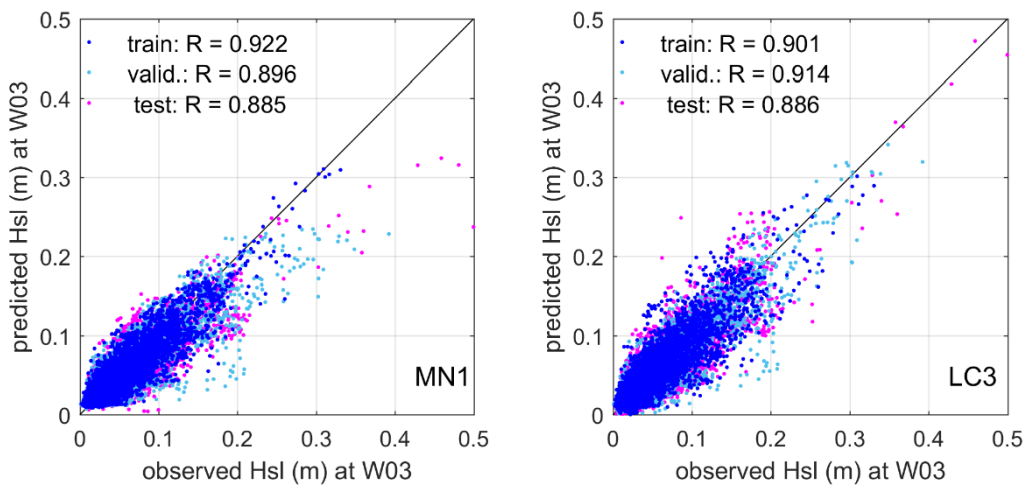
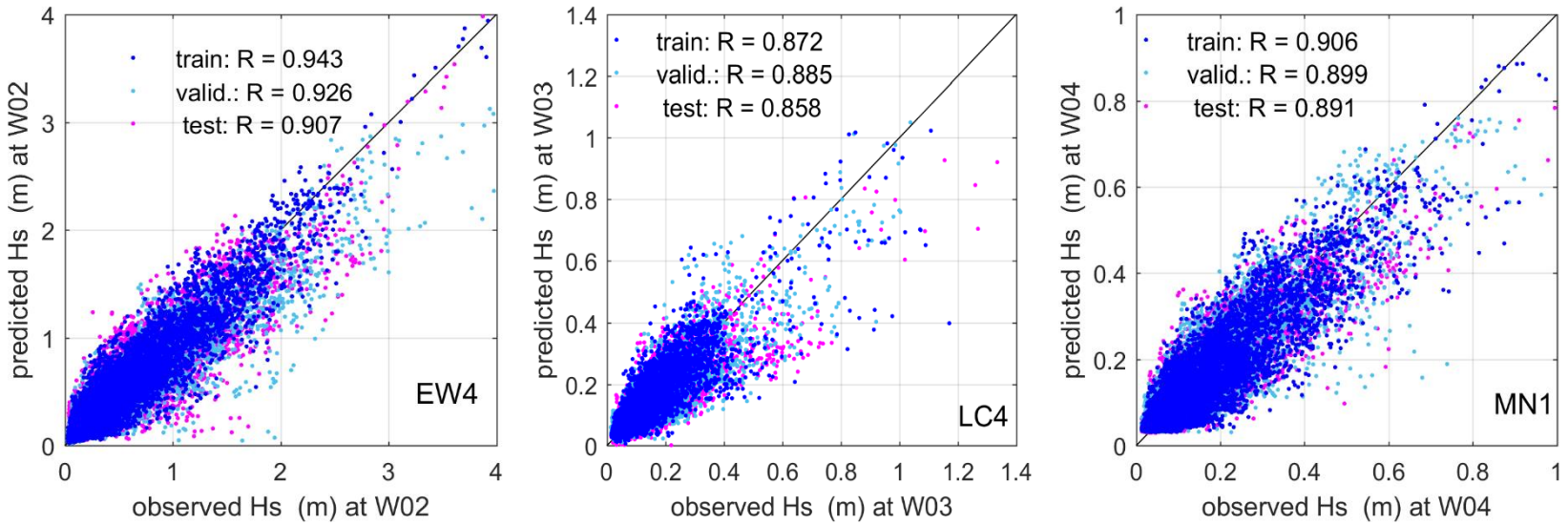


Figure 5.9 Performance comparison of MN1 and LC3 models for H_{sl} (W03)

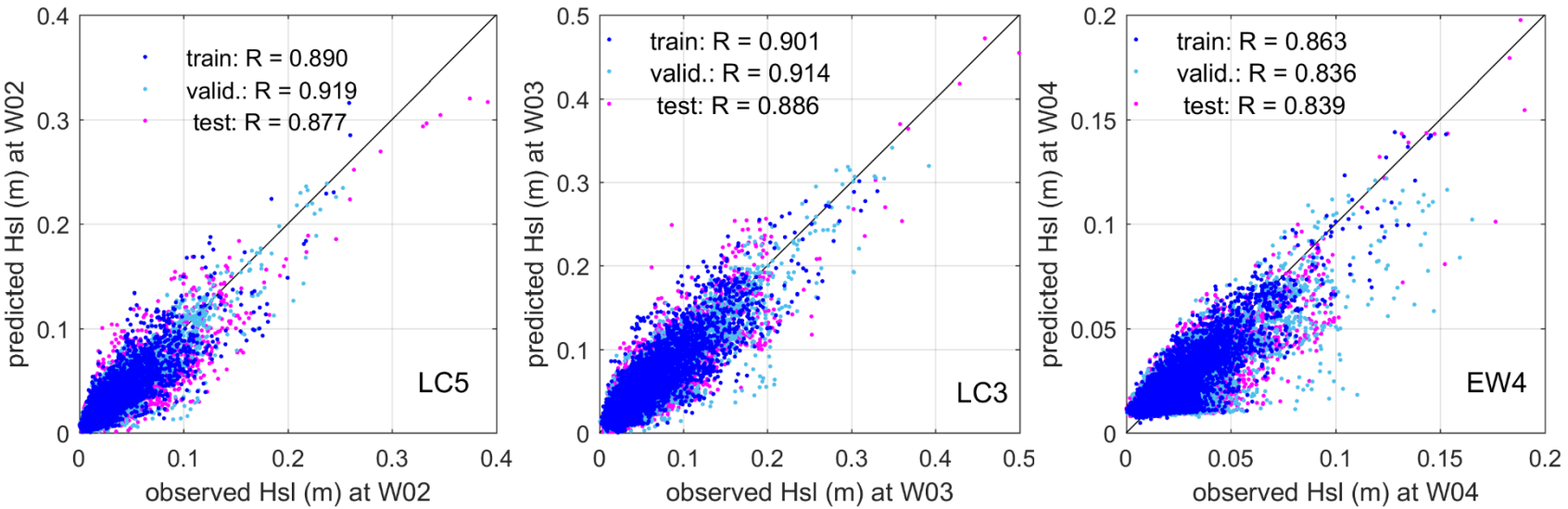
Table 5.3 Performance omparison of the different ensembles from the best selected model for each target variable

target	best model	ensemble	train NRMSE	test NRMSE	valid. NRMSE	train BIC	test BIC	valid. BIC	train R	test R	valid. R
Hs (W02)	**trainlm 10-[5 2]**	EW4	0.334	0.423	0.401	-3.676	-3.285	-3.068	0.943	0.907	0.926
Hs (W03)	**trainlm 10-4**	LC4	0.494	0.516	0.466	-6.099	-6.031	-5.957	0.872	0.858	0.885
Hs (W04)	**trainlm 10-[4 1]**	MN1	0.423	0.456	0.442	-5.948	-6.034	-5.777	0.906	0.891	0.899
Hsl (W02)	**trainlm 10-[3 1]**	LC5	0.460	0.480	0.394	-9.344	-8.995	-9.257	0.890	0.877	0.919
Hsl (W03)	**trainlm 10-7**	LC3	0.441	0.472	0.407	-8.330	-8.122	-8.147	0.901	0.886	0.914
Hsl (W04)	**trainlm 10-7**	EW4	0.506	0.557	0.567	-9.837	-9.422	-9.113	0.863	0.839	0.836
Tz (W02)	**trainbr 10-[9 1]**	MN1	0.500	0.516	0.577	-0.606	-0.568	-0.438	0.866	0.862	0.828
Tz (W03)	**trainbr 10-[5 1]**	MN1	0.528	0.566	0.590	-0.467	-0.323	-0.135	0.849	0.828	0.818
Tz (W04)	**trainbr 10-[8 1]**	MN1	0.542	0.554	0.550	-0.190	-0.400	-0.199	0.841	0.844	0.837



Hs (W02)	**trainlm**	**10-[5 2]**	**EW4**	**Hs (W03)**	**trainlm**	**10-4**	**LC4**	**Hs (W04)**	**trainlm**	**10-[4 1]**	**MN1**
	train	test	valid.		train	test	valid.		train	test	valid.
NRMSE	0.334	0.423	0.401	NRMSE	0.494	0.516	0.466	NRMSE	0.423	0.456	0.442
BIC	-3.676	-3.285	-3.068	BIC	-6.099	-6.031	-5.957	BIC	-5.948	-6.034	-5.777
R	0.943	0.907	0.926	R	0.872	0.858	0.885	R	0.906	0.891	0.899

Figure 5.10 The most effective ensemble model for the significant wave height of GW, H_s



Hsl (W02)	**trainlm**	**10-[3 1]**	**LC5**	**Hsl (W03)**	**trainlm**	**10-7**	**LC3**	**Hsl (W04)**	**trainlm**	**10-7**	**EW4**
	train	test	valid.		train	test	valid.		train	test	valid.
NRMSE	0.460	0.480	0.394	NRMSE	0.441	0.472	0.407	NRMSE	0.506	0.557	0.567
BIC	-9.344	-8.995	-9.257	BIC	-8.330	-8.122	-8.147	BIC	-9.837	-9.422	-9.113
R	0.890	0.877	0.919	R	0.901	0.886	0.914	R	0.863	0.839	0.836

Figure 5.11 The most effective ensemble model for significant wave height of IGW, H_{sl}

5.4 Discussion of the Prediction Models

A NN ensemble model for predicting waves at multiple stations inside and outside the Pohang New Port was constructed. Previous study investigated the relationship between IGW and GW by analyzing the observed wave data at Pohang New Port, and notable correlation was found between the GW and IGW among different wave stations. Considering this, one- and two-step prediction models were developed using the neural network and the nonlinear regression. The performance for some targets were reasonably good, while the result of some targets were poor, probably influenced by the complex layout of the port, shallow water depth, and nonlinear interaction of waves. For this reason, the NN prediction model shows better performance than the nonlinear regression model on average.

In addition to this, a number of one-step NN models were trained, and well configured models were picked through model selection strategy. Extending depth of NN model was effective for some targets. Furthermore, the ensemble method using NN results contributed to generalization of models. Instead of selecting the best individual network, building an ensemble network yielded simpler but more stable results for target variables.

In the vicinity of the port, the primary GW is quickly dissipated due to wave breaking although it is the dominant factor for occurring downtimes in the port. Hence, it is hard to precisely predict the magnitude of GW. On the other hand, IGW tends to be partially reflected on the shore rather than just be dissipated by breaking. The reflected IGW produces free long waves in the region, while the residual long waves are still correlated with the incident IGW.

Considering such complexities in wave transformation of GW and IGW, the predictability of GW and IGW by the NN prediction model in this study (the values of *R* close to 0.9) was quite good enough to be used as predictors of the downtime classification model. If such input variables are obtained by a fast-running NN model based on forecasted wave data provided by an operational wave model, the downtime forecasting can also be performed 24 to 72 hours earlier than a probable downtime event. Therefore, the operators and workers at the port will have sufficient time for coping with handling cargoes earlier before the probable downtime event occurs.

CHAPTER 6. CLASSIFICATION MODEL

6.1 Classification as a Downtime Forecasting Model

In order to secure tranquility of a harbor, the operation rate of the harbor is calculated in advance so as the rate to be higher than the acceptable level. The operation rate is dependent on the limit wave height for loading and unloading and the probability of wave height that exceeds the limit wave height. According to the design standards for a port (MOF, 2014), the limit wave height for long-period waves is just determined only considering the draft weight of a vessel. The limit for long period wave height is 0.2 m for vessels with the draft weight range of 1,000~5,000 ton. The limit is 0.15 m for that case of 5,000~10,000 ton and 0.10 m for that case of 10,000~70,000 ton. In fact, however, it is tough to discriminate whether a sea state is beyond a certain limit state or not when a downtime event occurs.

In this respect, maintaining harbor tranquility is a complicated task that cannot be secured by a universal threshold of limit wave height. Even though future wave height condition in a harbor is predicted and compared with the limit wave height, it does not provide real-time information on the actions for operation and management of the harbor to prevent downtime occurrences. Therefore, a classifier that forecasts downtime based on wave parameters supplied by a wave prediction model will be very helpful for port operators to take an effective measure in advance. If a well-trained classifier can

predict prospective downtime occurrence from the simulated data with high accuracy, the warning information of downtime can be provided tens of hours earlier (Figure 6.1). In this study, such a classifier was developed based on the classification method, which is a kind of machine-learning technology.

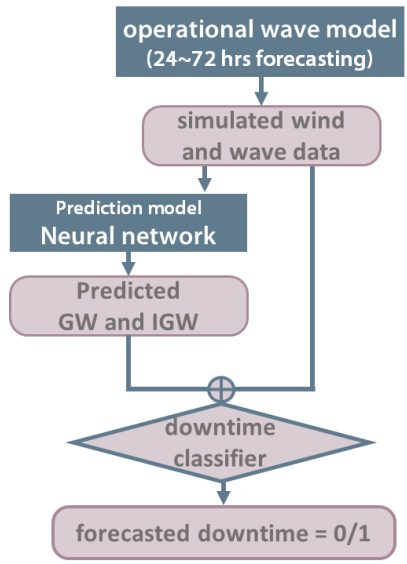


Figure 6.1 A flow chart of downtime forecasting model for practical application

6.2 Use of Predicted Wave and Modified Downtime Data

The classification method was applied to the wave data that are synchronized with the corresponding downtime data. The wave data to be used as input predictors of classification are the significant wave height of GW and IGW at W02, W03 and W04 stations, which are obtained by the NN prediction model described in Chapter 5. Meanwhile, the downtime data to be used as target responses of classification are binary-type data having values of either zero or one. As explained in Chapter 4, the downtime data was slightly modified by using the HHT analysis.

Exemplary scatter plots of showing the procedure of modifying the original downtime data is shown in Figure 6.2. As clearly shown in the figure, the scatter plot of the modified downtime data (panel (c) in Figure 6.2) presents more refined results after discarding and converting some portion of the data (panel (b) in Figure 6.2). The details of such a modification methodology have been elucidated in Section 4.3.2. A comparison of the original and modified downtime sequences is presented in Figure 6.3, where the solid area denotes the downtime occurrences while empty area denotes no downtime. As shown in the figure, a relatively small portions of the original downtime sequences was changed by the modification based on HHT analysis. In general, less than 10% of the original downtime data were modified.

Then, four different classification models were constructed to examine the effectiveness of the model by using either the observed or NN simulated wave data and either the original and modified downtime data. Figure 6.4 shows how the classification models are comprised depending on the use of data for predictor inputs and target responses. Each of the four classification models is termed as O1, O2, P1, and P2 as

illustrated in the figure. Based on some preliminary analysis, two different classification algorithms of k-Nearest Neighbors (kNN) and Quadratic Discriminant (Q.D.) were applied for each of the classification models because they are well fitted with the downtime classification model. For kNN algorithm, two different values of k=1 and 10 were used to examine the influence of the parameter. The details of the classification methodology and algorithms have been presented in Section 2.2.3.

Comparison of the results from the four classification models is summarized in Table 6.1. For both observed and predicted wave data, the use of modified downtime data apparently improved the performance of the downtime forecasting model. Irrespective of the three different classification algorithms, the model accuracy of O2 and P2 was higher than that of O1 and P1 respectively, demonstrating the superiority of the modified downtime data. Among the two classification algorithms investigated, kNN with either k=1 or k=10 showed slightly better performance than Q.D. Meanwhile, there was an only minor difference in model accuracy depending on the use of different k values for kNN algorithm.

Even though kNN algorithm provided superior classification results, Q.D. algorithm was also adopted for the downtime forecasting model because it provides more generalized and interpretable results of classification. In addition, the model performance of Q.D. algorithm was significantly improved when the modified downtime data were used for the classification model. The model accuracy when DT=1 increased by 5% with this algorithm, while kNN algorithm showed little change in precision. Figure 6.5 shows the scatter plots of the classification results with Q.D. algorithm. Considering the mixture of DT=0 and DT=1 data over a wide range of wave height, the accuracy of identifying downtime occurrences is extremely high and

acceptable for developing a downtime forecasting model.

Meanwhile, the model accuracy of the predicted wave data is slightly lower than that with the observed wave when DT=1. Comparison of P1 or P2 model results with those of O1 or O2 shows that the accuracy decreases by 7% with the former models. However, the model performance was comparable between the models with predicted and observed data when DT=0. Since the downtime forecasting model will be constructed based on the predicted wave data provided by an operational wave model, the eventual performance test will be carried out with P2 model in the following section.

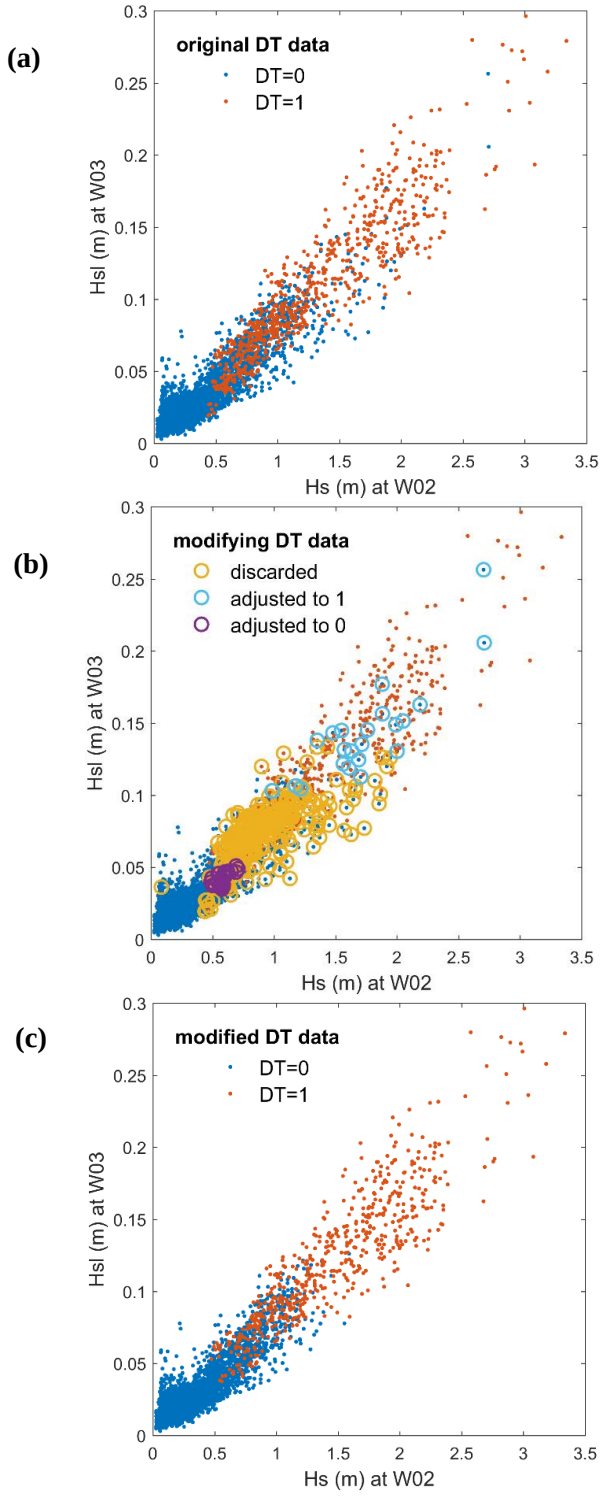


Figure 6.2 Scatter plots comparing the original and modified downtime data

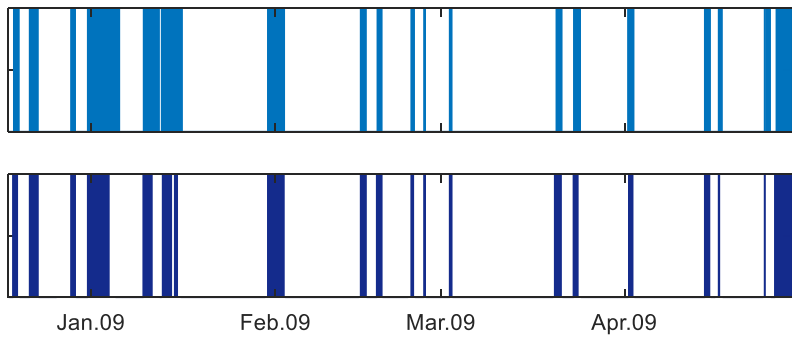


Figure 6.3 An exemplary time sequence of the original (upper) and modified (lower) downtime data use for the train set

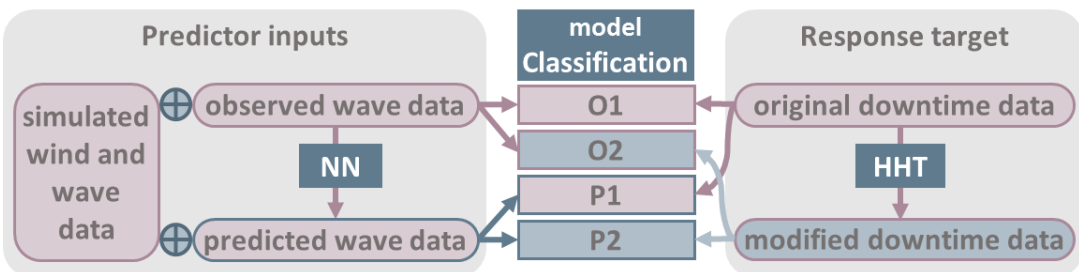
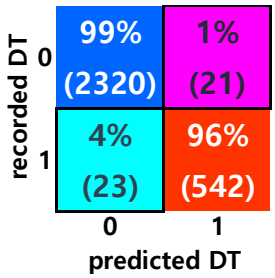
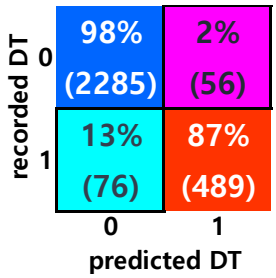
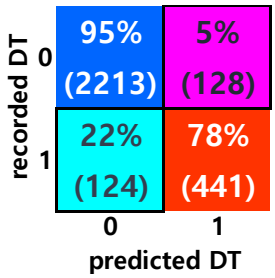
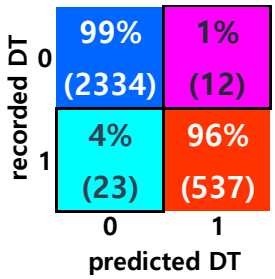
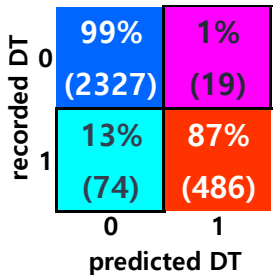
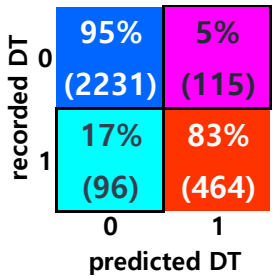
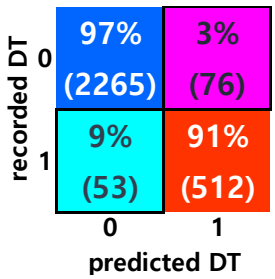
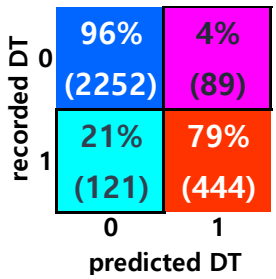
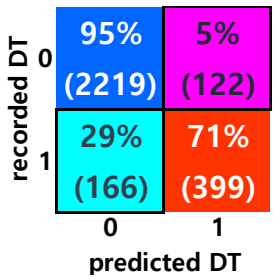
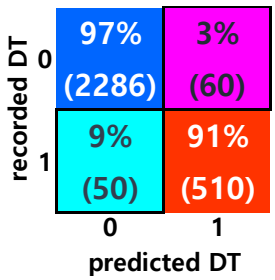
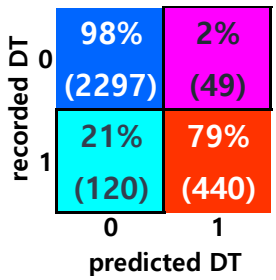
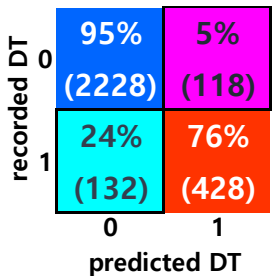


Figure 6.4 General diagram of the four classification models depending on predictor inputs and response targets

Table 6.1 Comparison of confusion matrices of model accuracy from the the four classification models with three different algorithms

model	kNN (k=1)	kNN (k=10)	Q.D.
O1			
O2			
P1			
P2			

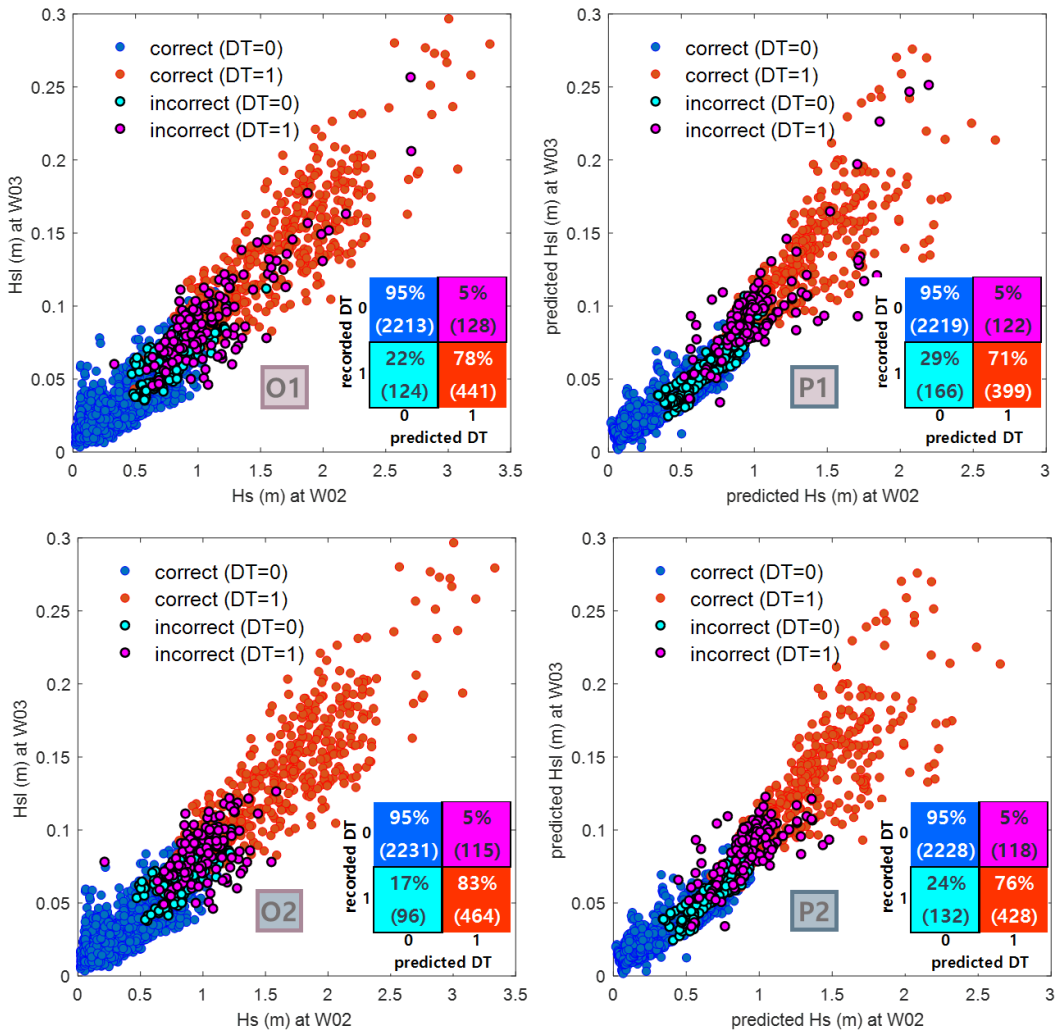


Figure 6.5 Scatter plots of the classification models with Q.D. algorithm

6.3 Test of Downtime Forecasting Model

A general overview of the practical use of the downtime forecasting model is shown in Figure 6.1. For such a practical application, it is required to test the effectiveness of the trained downtime classifier described in Section 6.2. For the concrete application of the downtime classifier, objects for downtime are specified and tested. For the test of downtime forecasting model, the quays around each wave observation station are identified by referring to Figure 3.3.; Quays 84 to 86 are for station the W03, quays 82 to 84 are for the station W04, and quays 75, 76, 81, 82 are for the station W10. The weight of vessels experiencing downtimes are investigated and categorized as illustrated in Figure 6.6. In the figure, distribution of the weight of the vessels which had downtime at each station is presented. For all the three groups of vessel weight, the downtime occurrences were the largest at W10. The most downtime occurred at W10 when there were relatively light-weight ships at the neighboring quays. It is considered that downtime was unavoidable even though the smaller ships tended to berth at inner side quays (around W10 station).

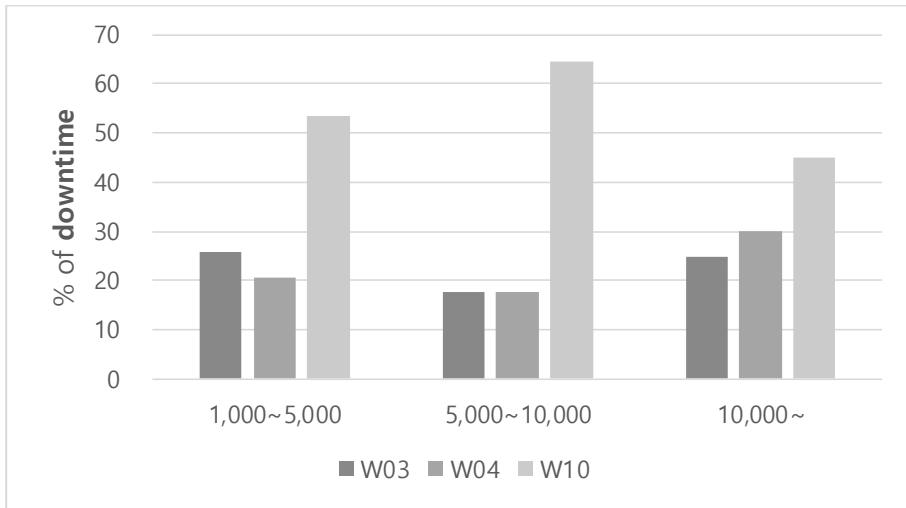


Figure 6.6 Histograms of the number of downtimes and relative percentages at the three wave stations for different groups of the vessel weight

The time sequences of downtime occurrence at each station presented in Figure 6.7 also shows that downtime is the most frequent at W10 station. When downtimes took place at W03 or W04 station, downtimes were always recorded at W10. The reason why the probability of downtime occurrence at W10 is higher than the other two stations can partly be influenced by the presence or absence of vessels at each location. If there were no ship at a certain location of the quay, it should be classified as no downtime (DT=0) even though high waves appeared at the location. Considering this, all the gathered downtime data of the three stations were used as target response of the classification model.

Then, the downtime sequences were presented according to the ranges of vessel weight as in Figure 6.8. Following the standard in harbor design regarding limit wave height of long-period waves (MOF, 2014), three different groups of vessel weight (1,000~5,000 ton, 5,000~10,000 ton and beyond 10,000 ton) were taken into account. The downtime sequences in the first two panels of Figure 6.8 are quite similar as the one in the last panel. This indicates that a vessel lighter than 5,000 ton will always experience downtimes if downtime events occur somewhere inside the port. In contrast, a vessel heavier than 5,000 ton may not have downtimes even though comparatively lighter vessels are suffering downtimes. In this context, it was required to conduct the classification analysis considering the groups of vessel weight.

The classifiers trained with downtime data between Dec.18, 2008 and Apr.28, 2009 were subjected to the test for the data between Aug.01, 2009 and Feb.10, 2010 taking into account of the groups of vessel weight. Predicted GW and IGW (H_s and H_{sl}) at W02, W03, and W04 from the NN model and simulated wind and wave data at C01 were utilized for input predictors of the classification. In addition, the modified

downtime data were assigned as the target responses. However, the downtime data used for the test set was not modified.

The classification result for the identifying downtime is summarized in Table 6.2, where confusion matrices for the three groups of vessel weight are shown. kNN method with one nearest neighbor (k=1) showed the best performance among the three algorithms tested, irrespective of the vessel weight, whereas kNN with ten nearest neighbors (k=10) and Quadratic Discriminant (Q.D.) showed lower performances. Considering intrinsic characteristics of the classification algorithms, kNN (k=1) may overfit the given data as it is a distance based algorithm. On the other hand, Q.D. creates a discriminant with the given data so that the method can be more advantageous in the generalization of classification although it showed relatively inferior performance in the training set.

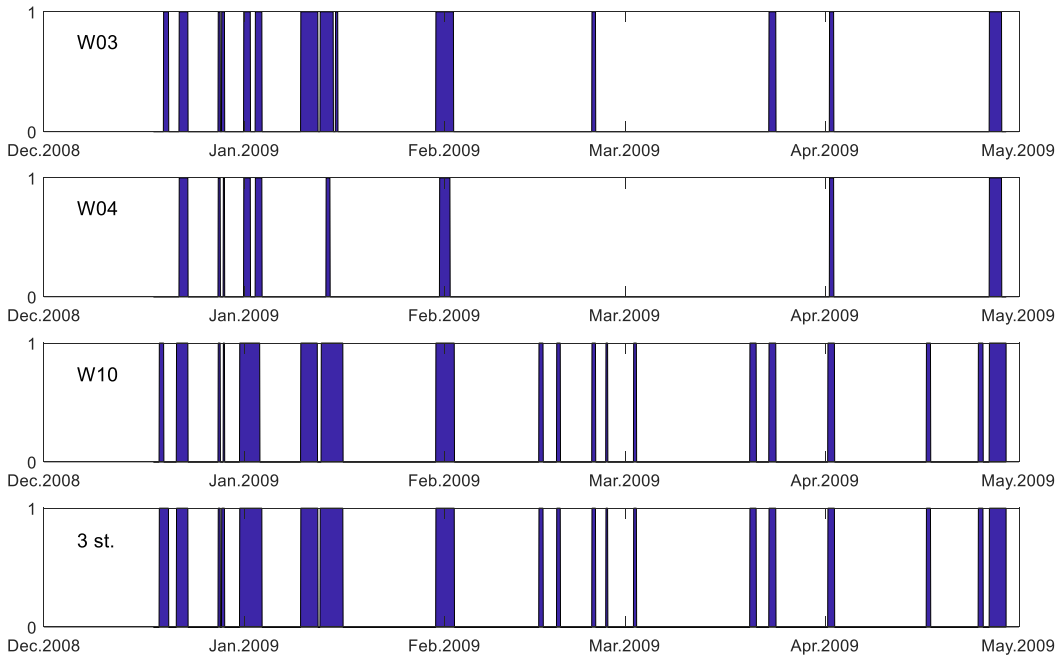


Figure 6.7 Downtime sequences at the three wave stations (W03, W04, and W10).

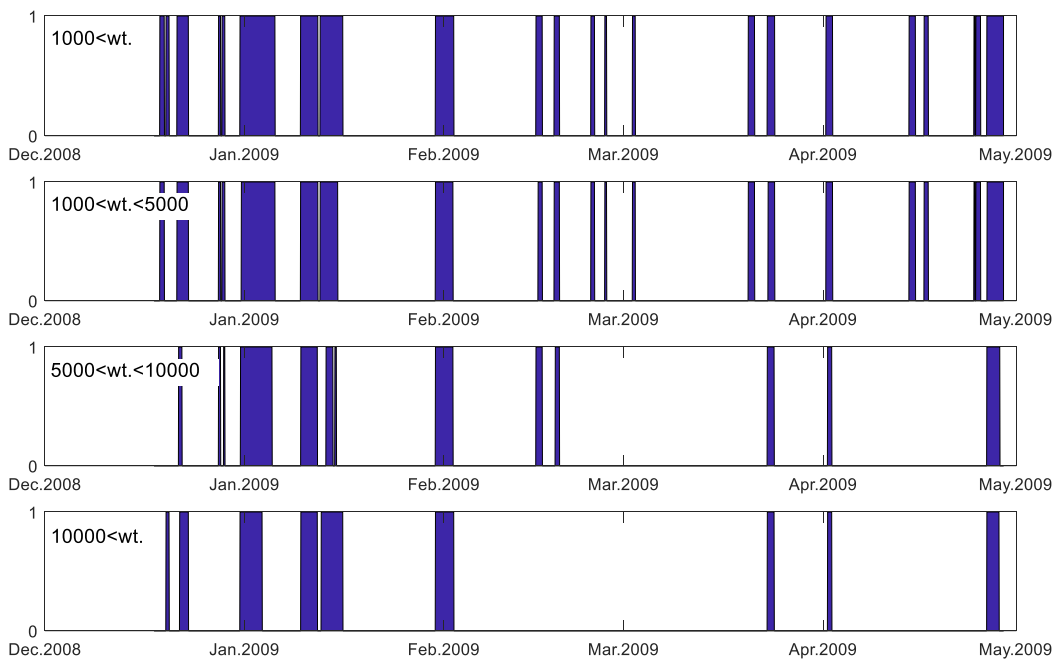
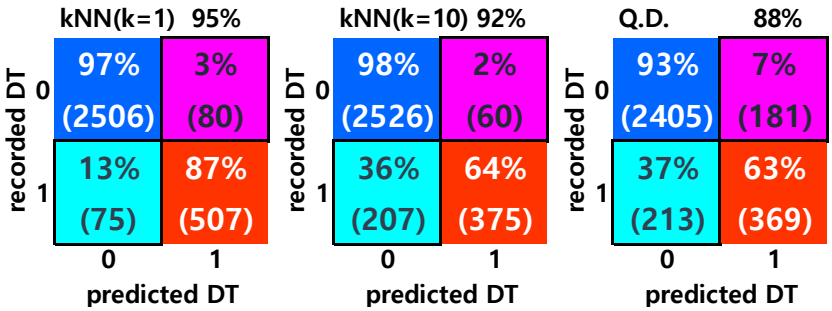


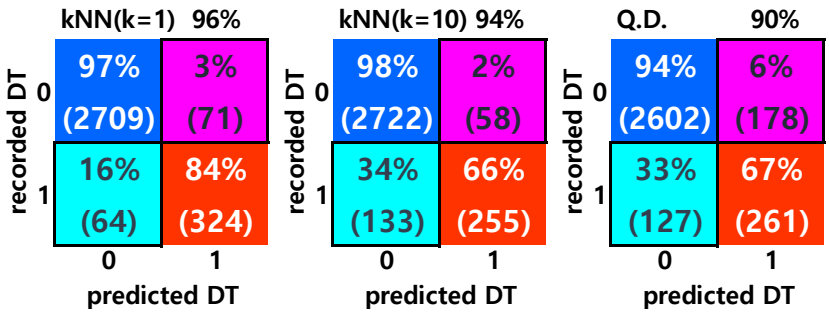
Figure 6.8 Downtime sequences with respect to the weight of ships

Table 6.2 Confusion matrices of the classification results

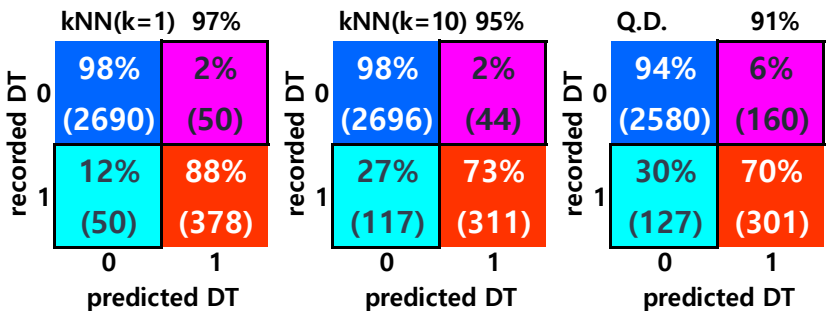
1,000 ton < Weight of ship <5,000 ton



5,000 ton < Weight of ship <10,000 ton



10,000 ton < Weight of ship



The downtime response was obtained by inputting the predictor of test data set into the trained classifier for vessel weight of 1,000~5,000 ton in Figure 6.9. Although kNN (k=1) showed the best results in training, it overpredicted the recorded DT in the test set as shown in the figure. Similar results were obtained with kNN (k=10) and Q.D. However, if we only keep the downtimes commonly predicted by all the three algorithms, the resulting downtime sequence, termed as predicted DT in the figure, shows a very similar sequence to the recorded downtime (DT). Therefore, the predicted DT obtained by a combination of the three classifiers can be satisfactorily used to predict the prospective downtime events.

Figure 6.10 presents scatter plot and confusion matrix between the predicted DT and the recorded DT. When the predicted DT is the same as the recorded DT, the case is marked as correct (DT=0/1). When the classifier gives a wrong prediction, the case is marked as incorrect (DT=0/1). The percentage written in the confusion matrix represents the accuracy of the prediction for the each case, while the percentage (92%) on the top of the matrix means the total percentage of correctly classified DT. The accuracy of prediction was 95% when there was no downtime, but the accuracy was much lower as 59% when downtimes occurred. Most of the incorrectly predicted DT=0 (cyan: 41%) is related to relatively smaller wave height. Except that, the overall performance of the downtime classifier was quite effective, coinciding well with the recorded DT.

The test results for vessels of 5,000~10,000 ton are shown in Figure 6.11 and 6.12. In this case, downtime sequence of the predicted DT showed significant difference from the recorded DT, although the classifier represents high performance (94%) in the percentage of correctly classified DT. Indeed, only 8% of the downtime records were

correctly predicted when downtime event occurs. However the total number of downtime events belong to this group was only five (see the number in the parenthesis in the confusion matrix). Most of the incorrectly predicted DT=0 (cyan: 92%) events were associated with relatively smaller waves. In contrast, most of the incorrectly predicted DT=1 (magenta: 5%) events were associated with relatively higher waves. Consequently, the predicted downtimes may not fit well with the recorded downtimes for the vessels in this weight ranges, especially when the wave height is comparatively small.

Lastly, the test results for vessels heavier than 10,000 ton are shown in Figure 6.13 and 6.14. The accuracy of prediction was 90% when there was no downtime, whereas 67% when downtime occurred. The total percentage of correctly classified DT is 89%. The coincidence between the downtime sequences of predicted DT and recorded DT was much lower than the results for vessels of 1,000~5,000 ton, displaying over-prediction of downtime occurrence.

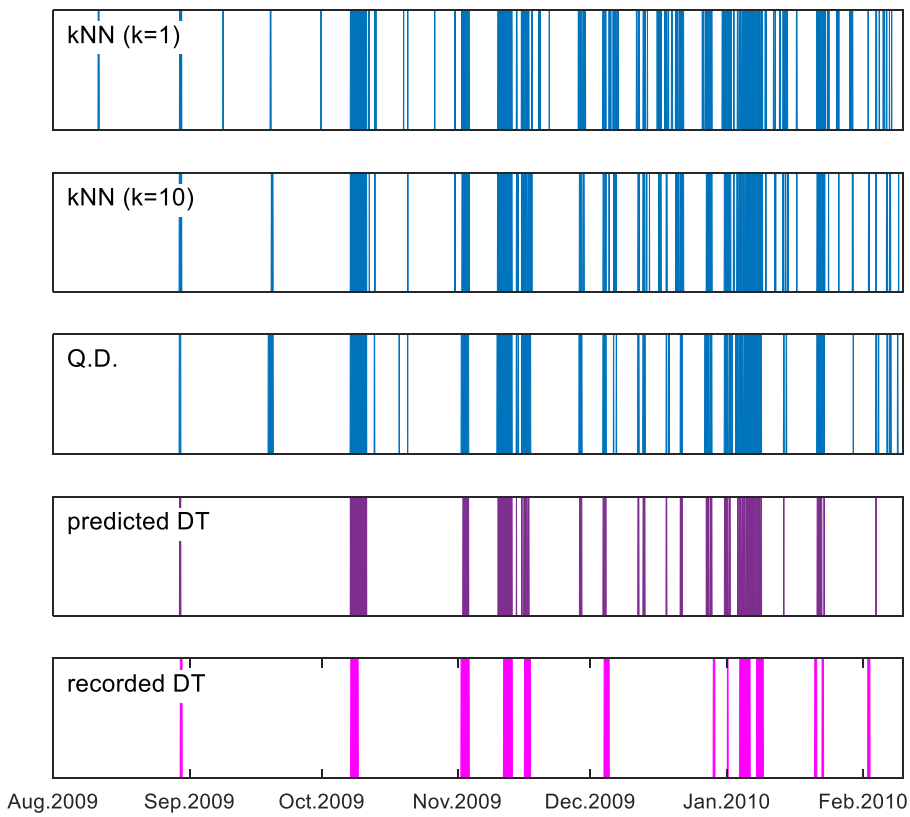


Figure 6.9 Comparison of the predicted and recorded downtime sequences for the vessel weight between 1,000 and 5,000 ton

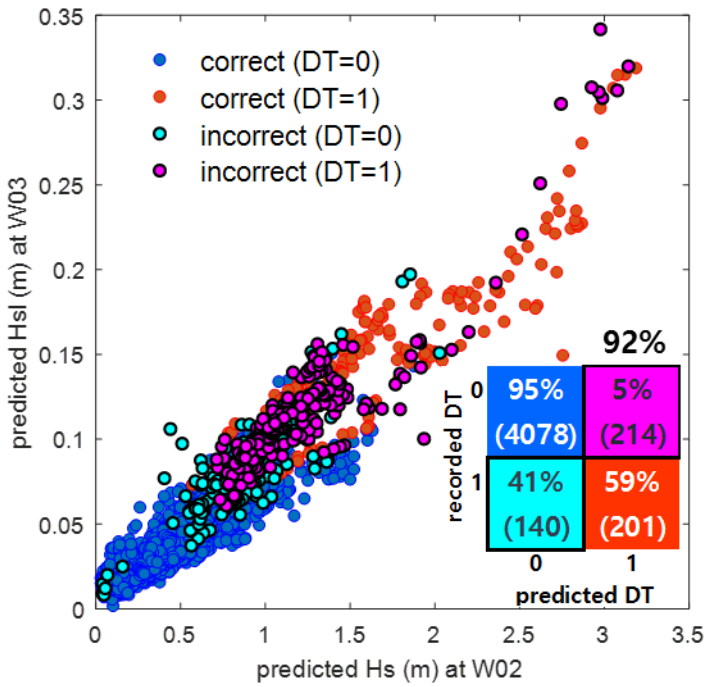


Figure 6.10 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight between 1,000 and 5,000 ton

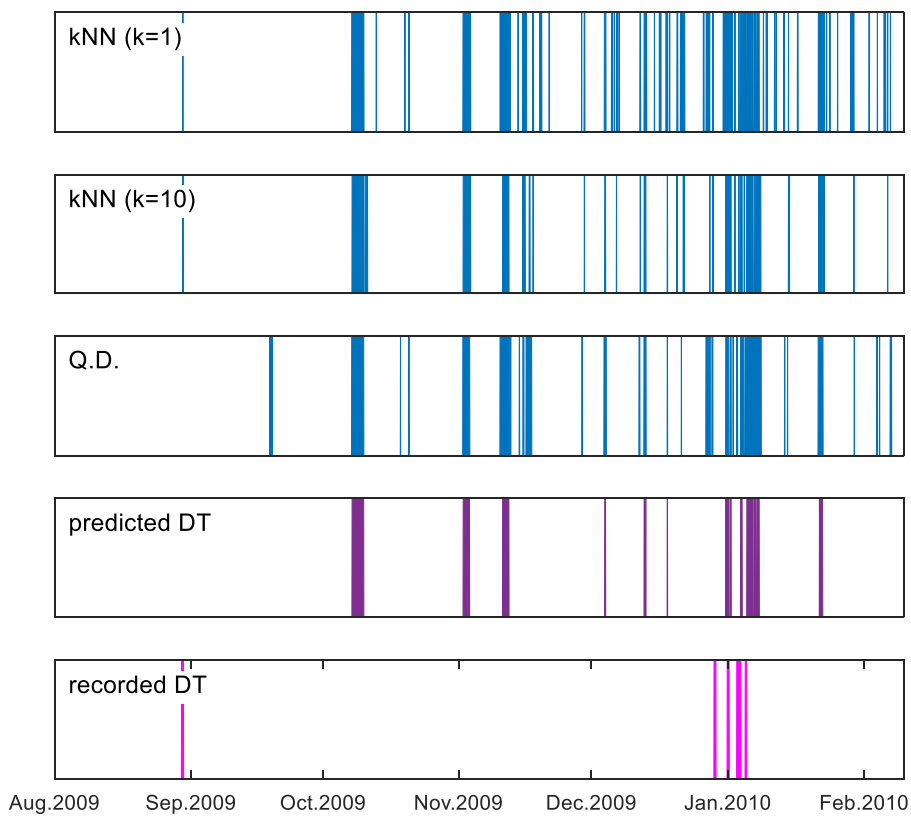


Figure 6.11 Comparison of the predicted and recorded downtime sequences for vessel weight between 5,000 and 10,000 ton

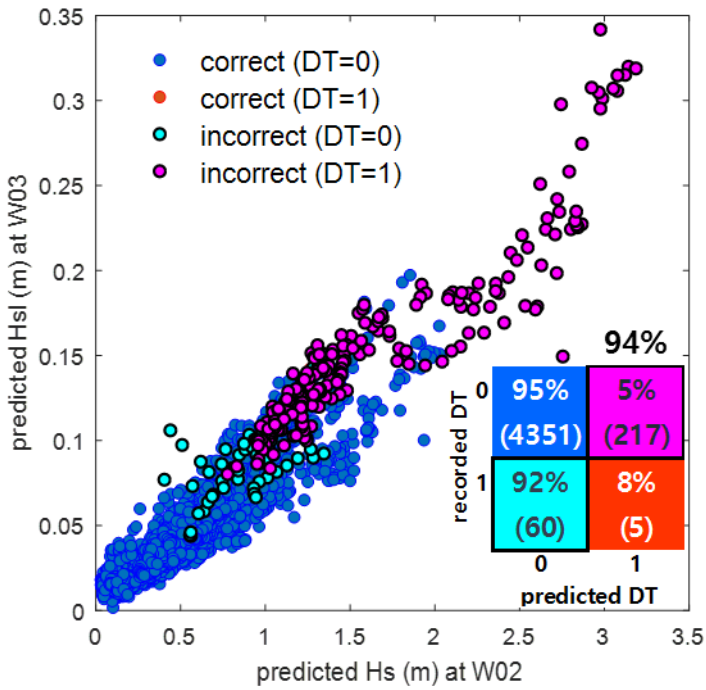


Figure 6.12 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight between 5,000 and 10,000 ton

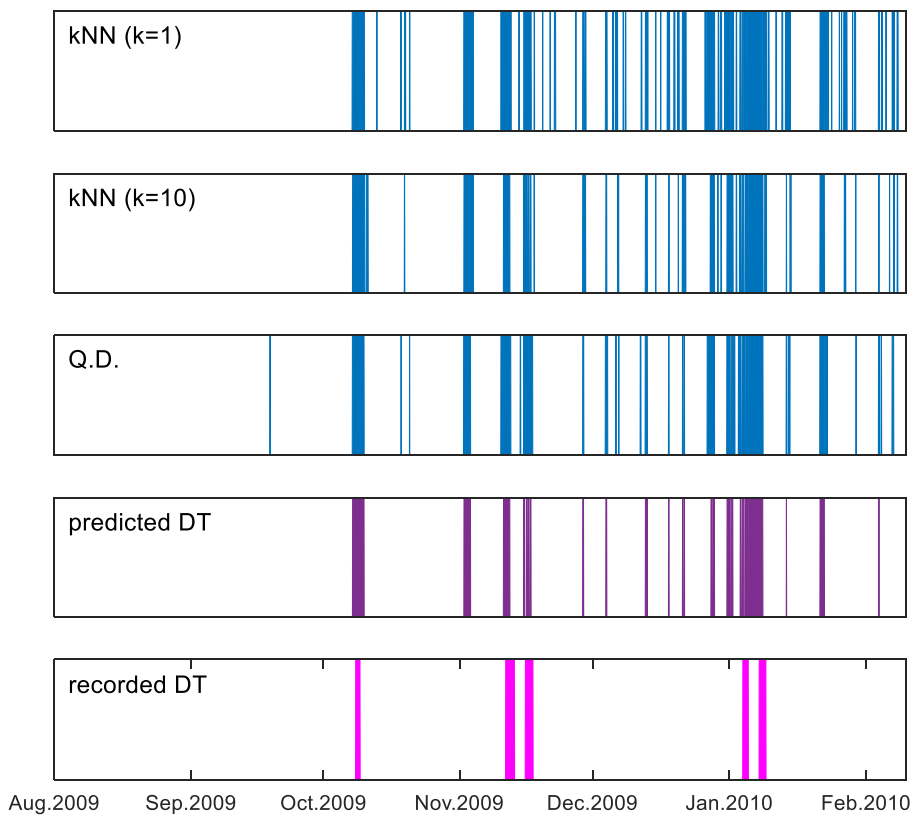


Figure 6.13 Comparison of the predicted and recorded downtime sequences for vessel weight heavier than 10,000 ton

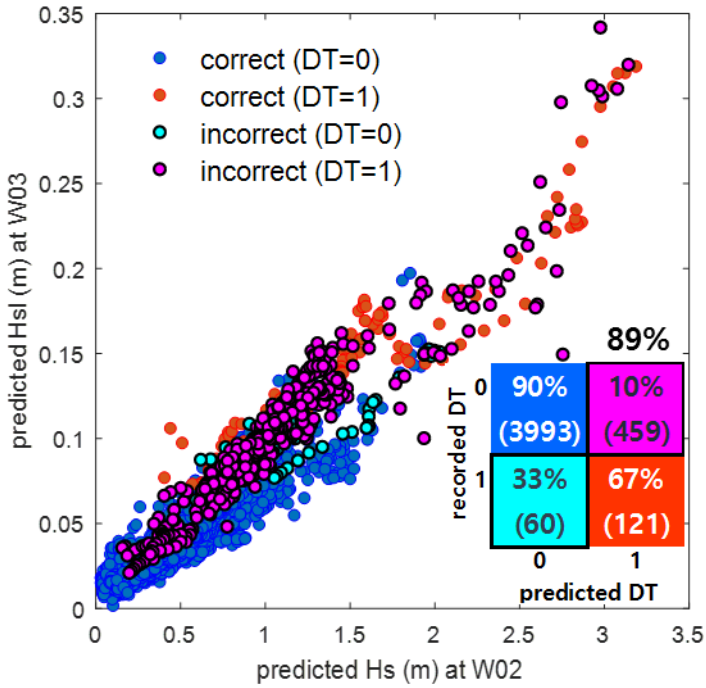


Figure 6.14 Scatter plot and corresponding confusion matrix of the predicted downtime for vessel weight heavier than 10,000 ton

As described in the above, the downtime forecasting results for vessels of 1,000~5,000 ton, which is the representative vessel weight range in the port, was found to be highly reliable. On the contrary, the predictive capability became poorer with the increase of vessel weight, showing a tendency of over-predicting downtimes. The poor performance of downtime forecasting model for relatively heavier vessels is partly due to insufficient sample size (or data points) to train the classifier, and partly because the used input predictors were not suitable for downtime prediction for the those vessel weight ranges. Another possible reason can be use of the original (unmodified) downtime data in the test set.

In this context, a sufficient number of data sets for heavier vessels are required to produce a better classifier and to build a more reliable forecasting model that can be effectively used over the whole ranges of the vessel weight. The input features of the classification for heavier vessels can be changed for more probable classification for that range of vessels. In addition, the performance of the downtime forecasting model is expected to be improved if more downtime data sets are obtained at the quays of the port.

CHAPTER 7. CONCLUSION

Harbors with complex layouts are prone to experience severe agitation and downtime problems. The cause of such problem was hard to be revealed and prediction of downtime was not straightforward so far. In this context, this study try to enhance understanding of the influence of shorter- and longer-period waves on the agitation in the port and to develop a forecasting model identifying downtime events at the quays in the Pohang New Port. The downtime forecasting model was developed based on data modeling methods by using wave and meteorological data available around the harbor. For constructing a data-based forecasting model, extensive wave monitoring and simulation data together with manually-written downtime records were collected and compiled to be used for subsequent analyses.

At the initial stage of the analysis, the obtained downtime records were transformed into a binary-type hourly data that were used as target responses of a classification method. By doing this, a subset of the downtime data can be extracted according to some conditional properties (*e.g.* quay location, vessel weight and so on) that are required in the analysis. Hence, it can be used as a target object for a specific classification model that can be used under the pre-defined conditional properties. In the preliminary analysis of the downtime data, the overall coincidence between high waves and downtime occurrences was recognized, but a specific trend or condition for downtime occurrence could not be figured out.

The relationships between downtime occurrence and spectral energy increase in wave

components of different periods were investigated by using HHT analysis. As a result, it was confirmed that GW components are the primary factor for harbor agitation and subsequent downtime occurrences, as already pointed out some previous researches. Moreover, it was found that IGW components also have an apparent secondary influence on the downtime, which was not disclosed apparently in the previous researches. Also, it was demonstrated that very long period agitations related to natural oscillation period (NOP) of Pohang New Port and Youngil bay have little association with the downtime. Based on the time variation of energy in the HHT spectrum, some erroneous downtime data were eliminated or modified, and more refined downtime data were prepared to be used for the forecasting model.

For preparing input variables for the downtime forecasting model, nonlinear regression models and NN models were adopted to predict wave parameters by using available wind and wave data. Although the both models showed satisfactory performance in predicting the wave parameters, NN model provided slightly better results with almost comparable computing speed, resulting in a selection of NN as a fundamental methodology for obtaining necessary wave parameters as input predictors of the forecasting model. The NN model can predict wave parameters of both the shorter-period GW and longer-period IGW inside and outside the port by using numerically simulated data. By comparing the model performance with different configurations of the parameters, the best NN model for each parameter was obtained. Then, ensemble operations with the best NN model were carried out for more generalized results. The correlation coefficient, R was close to 0.9 for all the predicted parameters, which demonstrates that the predicted results of the model was good enough to be used as input variables of the downtime forecasting model.

By using all the above data and model outputs, a classification model for the downtime in Pohang New Port was developed. The model results in training set show that the classification method developed in this study well classifies whether a downtime occurs or not, implying the appropriateness of selecting all the parameters used by the model. Furthermore, the model results with the modified downtime data showed better performance than those with the original downtime data, confirming the effectiveness of the modification of the downtime data by using HHT analysis.

Then, the classification model was subjected to a series of test to examine whether the model could be eventually used for practical downtime forecasting in real-time operational model. The test was conducted with different groups of the data in terms of draft weight of vessels. The forecasting result for the vessels of 1,000 to 5,000 tons showed fairly good agreement between the predicted downtime and the recorded downtime. More than 80% of the recorded downtime was well forecasted by the classification model. Hence, the downtime forecast may provide a practical warning information on when prospective downtime occurs and terminates. Meanwhile, the forecasting results for large vessels weighing over 5,000 ton showed comparatively inferior performance. It is necessary to investigate whether such a result is due to small number of samples used for developing the classification model or imperfect feature selection of the classification model.

Sufficient length of data for constructing a data-based model is varying depending on target problems and modeling methods. The NN model used in this study to predict wave parameters needs training period of at least two winter seasons. The wave field on the Korean east coast becomes more severe in winter season, so inclusion of data during winter season is crucial for developing the model. Meanwhile, the training period longer

than two winter seasons is also required to make an efficient classifier for downtime event. In this study, however, the training period of four and half months was used and it was not fully sufficient to develop a good model. More validated forecasting model can be obtained if length of the training period satisfies the aforementioned standard.

Nowadays operational wave prediction models are providing forecasted wave and wind data earlier than tens of hours. Hence, the data-based downtime forecasting model developed in this study can also provide an information of prospective downtime for upcoming hours comparable to the wave prediction models because the model runs very fast and it is straightforward to implement into the model. If the downtime forecasting model is tested and validated with more downtime cases and corresponding data, the performance will be more improved by the intrinsic attribute of machine-learning technology. In addition, the downtime forecasting model can be extended to other ports in Korea or around the world where downtime problems are occurring provided that sufficient data for constructing the model are available. In this respect, practical application of the data-based downtime forecasting model in coming years is very promising and it will contribute to enhance safety of the port operation by preventing possible damage of vessels or harbor infrastructure due to unexpected downtimes.

References

Adhikari, R., 2015. A neural network based linear ensemble framework for time series forecasting. *Neurocomputing*, 157, 231-242.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.

Armstrong, J.S., 2001. *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer Science & Business Media, 850p.

Bellotti, G. and Franco, L., 2011. Measurement of long waves at the harbor of Marina di Carrara, Italy. *Ocean Dynamics*, 61(12), 2051-2059.

Bowers, E.C., 1992. Low frequency waves in intermediate water depths. *Proc. 23rd International Conference in Coastal Engineering*, ASCE, 832-845.

Chen, J.; Xu, Y.L. and Zhang, R.C., 2004. Modal parameter identification of Tsing Ma suspension bridge under Typhoon Victor: EMD-HT method. *Journal of Wind Engineering & Industrial Aerodynamics*, 92, 805-827.

Cho, Y.J.; Cho, Y.H., and Kim, M.K., 2010. Seiche characteristics of the Pohang Young-il Bay centered on edge waves. Proceedings of Joint Conference of Korea Association of Ocean Science and Technology, 2697-2701 (in Korean).

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4), 559-583.

Deka, P.C. and Prahlada, R., 2012. Discrete wavelet neural network approach in

significant wave height forecasting for multistep lead time. *Ocean Engineering*, 43(January), 32-42.

Dixit, P. and Londhe, S., 2016. Prediction of extreme wave heights using neuro wavelet technique. *Applied Ocean Research*, 58, 241-252.

Eğrioğlu, E.; Aladağ, C.H., and Günay, S., 2008. A new model selection strategy in artificial neural networks. *Applied Mathematics and Computation*, 195(2), 591-597.

Elgar, S. and Guza, R.T., 1985. Observations of bispectra of shoaling surface gravity waves. *Journal of Fluid Mechanics*, 161, 425-448.

Elgar, S.; Herbers, T.H.C.; Okihiro, M.; Oltman-Shay, J., and Guza, R.T., 1992. Observations of infragravity waves. *Journal of Geophysical Research*, 97 (C10), 15,573-15,577.

Goda, Y., 2010. *Random seas and design of maritime structures*. World scientific, 708p.

Granger, C.W. and Ramanathan, R., 1984. Improved methods of combining forecasts. *Journal of forecasting*, 3(2), 197-204.

Grouven, U.; Bergel, F., and Schultz, A., 1996. Implementation of linear and quadratic discriminant analysis incorporating costs of misclassification. *Computer methods and programs in biomedicine*, 49(1), 55-60.

Guza, R.T. and Thornton, E.B., 1982. Swash oscillations on a natural beach. *Journal of Geophysical Research: Oceans*, 87(C1), 483-491.

Guza, R.T. and Thornton, E.B., 1985. Observations of surf beat. *Journal of Geophysical Research: Oceans*, 90(C2), 3161-3172.

Hansen, L.K. and Salamon, P., 1990. Neural network ensembles. *IEEE transactions*

on pattern analysis and machine intelligence, 12(10), 993-1001.

Herbers, T.H.C.; Elgar, S., and Guza, R.T., 1994. Infragravity-frequency (0.005-0.05 Hz) motions on the shelf. Part I: Forced waves. *J. Phys. Oceanogr*, 24 (5), 917-927.

Herbers, T.H.C.; Elgar, S., and Guza, R., 1995. Generation and propagation of infragravity waves. *Journal of Geophysical Research*, 100(12), 863-872.

Huang, N.E.; Long, S.R., and Shen, Z., 1996. The mechanism for frequency downshift in nonlinear wave evolution. *Adv. Appl. Mech*. 32, 59-111.

Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; ... and Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1971), pp. 903-995.

Huang, N.E.; Shen, Z., and Long, S.R., 1999. A new view of nonlinear water waves: The Hilbert Spectrum 1. *Annual Review of Fluid Mechanics*, 31(1), pp.417-457.

Huang, N.E.; Chern, C.C.; Huang, K.; Salvino, L.W.; Long, S.R., and Fan, K.L., 2001. A new spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. *Bulletin of the Seismo- logical of America*. 91(5), 1310-1338.

Huntley, D.A., 1976. Long-period waves on a natural beach. *Journal of Geophysical Research*, 81(36), 6441-6449.

Huntley, D.A.; Guza, R.T., and Thornton, E.B., 1981. Field observations of surf beat: 1. Progressive edge waves. *Journal of Geophysical Research: Oceans*,

86(C7), 6451-6466.

Hwang, P.A.; Kaihatu, J.M., and Wang, D.W., 2002. A comparison of the energy flux computation of shoaling waves using Hilbert and wavelet spectral analysis techniques. *7th Int. Workshop on Wave Hindcasting and Forecasting*. 10, pp. 21-25.

Jeong, W.M.; Ryu, K.H.; Baek, W.D., and Choi, H.J., 2011. Downtime analysis for Pohang new harbor through long-term investigation of waves and winds. *Journal of Korean Society of Coastal and Ocean Engineers*, 23(3), 226-235 (in Korean).

Jeong, W.M.; Oh, S.H., and Eum, H.S., 2016. Analysis of Wave Climate around Korea Based on Long-Term Hindcast and Coastal Observation Data, *Journal of Coastal Research*, SI75, 735-739.

Jeong, W.M., 2017. Wave Hindcast Modeling around Youngil-Bay Using JMA and SWAN, personal communications.

Kankal, M. and Yüksek, Ö., 2012. Artificial neural network approach for assessing harbor tranquility: The case of Trabzon Yacht Harbor, Turkey. *Applied Ocean Research*, 38, 23-31.

Kinsman, B., 1965. *Wind waves: their generation and propagation on the ocean surface*. Courier Corporation, 676p.

Korgen, B.J., 1995. Transient standing-wave oscillations in water bodies can create hazards to navigation and unexpected changes in water conditions. *American Scientist*, 83(4), 330-341.

Kumar, P.; Zhang, H., and Kim, K.I., 2014. Spectral Density Analysis for Wave Characteristics in Pohang New Harbor. *Pure and Applied Geophysics*,

171(7), 1169-1185.

Longuet-Higgins, M.S. and Stewart, R.W., 1962. Radiation stress and mass transport in gravity waves, with application to “surf beats.” *Journal of Fluid Mechanics*, 13(4), 481-504.

López, M. and Iglesias, G., 2014. Long wave effects on a vessel at berth. *Applied Ocean Research*, 47, 63-72.

Mei, C.C. and Agnon, Y., 1989. Long-period oscillations in a harbour induced by incident short waves. *Journal of Fluid Mechanics*, 208, 595-608.

Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B., and Mullers, K.R., 1999. Fisher discriminant analysis with kernels. In Neural Networks for Signal Processing IX, 1999. *Proceedings of the 1999 IEEE Signal Processing Society Workshop*. pp. 41-48.

Ministry of Oceans and Fisheries, 2014. *Design Standards for Port and Fishing Port* (in Korean).

Mitchell, T.M., 1997. *Machine Learning*. McGraw Hill. 45(37), pp.870-877.

Nagai, T.; Hashimoto, N.; Shiraishi, S.; Kawaguchi, K.; Shimizu, K.; Ueno, S.; Kobayashi, A.; Toue, T., and Shibata T., 1997. Continuous Observations on Development and Decay of Offshore Long Waves and Long Period Oscillations of Moored Vessels, *Proc. Coastal Engineering, JSCE*, 44, 226-230.

Okihiro, M. and Guza, R.T., 1996. Observations of seiche forcing and amplication in three small harbors. *Journal of waterway, port, coastal, and ocean engineering*, 122(5), 232-238.

Peng, Z.K.; Tse, P.W, and Chu F.L., 2005. A comparison study of improved Hilbert-

Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, 19(5), 974-988.

PIANC, 1995. Criteria for movements of moored ships in harbours: a practical guide, report of working group no. 24 of the Permanent Technical Committee II. Brussels, Belgium.

Pohang Regional Maritime Affairs and Port Office, 2010. F*ield Investigation Report for Swell Analysis and Improvement Measure Making of Downtime in Pohang New Harbor*, Report of Pohang Regional Maritime Affairs and Port Office (in Korean).

Poularikas, A.D. (ed.), 1998. *Handbook of formulas and tables for signal processing* (Vol. 13). CRC Press, 864p.

Quek, S.T.; Tua, P.S., and Wang, Q., 2003. Detecting anomalies in beams and plate based on Hilbert- Huang transform of real signals. *Smart Materials and Structures*, 12, 447-460.

Rabinovich, A.B., 2009. *Seiches and Harbour Oscillations*. Handbook of Coastal and Ocean Engineering, pp.193-236.

Ruessink, B.G., 1998. Bound and free infragravity waves in the nearshore zone under breaking and nonbreaking conditions. *Journal of Geophysical Research*, 103(C6), 12,795-12,805.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.

Shahabi, S.; Khanjani, M.J., and Kermani, M.R.H., 2017. Significant wave height modelling using a hybrid Wavelet-genetic Programming approach.

KSCE Journal of Civil Engineering, 21(1), 1-10.

Spanish Ministry of Public Works, 2000. ROM 3.1-99, proyecto de la configuración marítima de los puertos canales de acceso y áreas de flotación. Madrid, Spain.

Suhayda, J.N., 1974. Standing waves on beaches. *Journal of Geophysical Research*, 72(21), 3065-3071.

Tan, P.N.; Steinbach, M., and Kumar, V., 2006. *Introduction to data mining*. Pearson Addison-Wesley, 736p.

van der Molen, W., 2006. *Behaviour of moored ships in harbours*. (Doctoral dissertation, TU Delft, Delft University of Technology).

Veltcheva, A. and Guedes Soares, C., 2016. Analysis of Wave groups by the Hilbert Huang Transform method. *Applied Ocean Research*, 60, 176-184.

Veltcheva, A. and Guedes Soares, C., 2016. Nonlinearity of abnormal waves by the Hilbert-Huang Transform method. *Ocean Engineering*, 115, 30-38.

Veltcheva, A., 2002. Wave and group transformation by a Hilbert spectrum. *Coastal Engineering Journal*, 44(4), 283-300.

Willmott, C.J., 1981. On the validation of models. *Physical Geography*, 2(2), 184-194.

Wright, L.D.; Chappell, J.; Thom, B.G.; Bradshaw, M.P., and Cowell, P., 1979. Morphodynamics of reflective and dissipative beach and inshore systems: Southeastern Australia. *Marine Geology*, 32(1-2), 105-140.

Wu, J.K. and Liu, P.L.F., 2006. Harbour excitations by incident wave groups. *Journal of Fluid Mechanics*, 217(1), 595-613.

Wu, X.; Kumar, V.; Quinlan, J. R.; Ghosh, J.; Yang, Q.; Motoda, H.; ... and Zhou,

Z.H., 2008. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1), 1-37.

Zhou, Z.H.; Wu, J., and Tang, W., 2002. Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1-2), 239-263.

국문초록

하역중단 예보를 위한 자료기반 모델: 포항신항에 적용

오정은

건설환경공학부

서울대학교

포항신항과 같은 복잡한 형상의 항만에서 예기치 않게 발생하는 하역중단은 항만운용에 있어 매우 중대한 문제이다. 이러한 하역중단은 항만의 현장관측이나 수치수치모델링부터 잘 예측되지 못했다. 따라서, 이러한 하역중단에 대한 내재적 원인에 대한 이해가 더 필요할 뿐 아니라, 항만운용을 위한 실제적 도움이 되는 하역중단에 대한 예보 정보를 구축하는 노력도 필요하다.

종래의 수치모델링으로 항만 가동률을 계산하여 항만정온도를 평가했던 것과 달리, 본 연구에서는 포항신항 인근의 대규모 파랑관측자료와 수집된 하역중단 기록을 바탕으로 하역중단에 대한 자료기반 예측모델을 구축하였다. 예측모델의 주요 구조는 항만 안팎의 중력파와 중력외파의 파랑변수들을 파랑추산된 인근의 파 바람 자료를 사용하여 인공신경망으로부터 추정하는 것으로 시작된다. 추정된 항만의 파랑변수들은 하역중단 자료를 목표변수로 하는 분류모델의 입력변수로 활용되는데, 본 연구에서 인공신경망에 의한 파랑변수의 예측성능은 90%에 근접하며 분류모델에서 하역중단 예측

을 위한 입력값으로 활용되는데 적절한 것으로 평가된다.

이에 더하여 힐버트-후앙 트랜스폼으로 하역중단 상황에서의 다양한 파랑성분들의 스펙트럼에너지에 대한 시공간적 변화를 관찰할 수 있었다. 이를 통하여 중력파와 중력외파 및 항만에서의 공진주기 등이 하역중단에 미치는 영향이 고찰되었다. 이러한 분석은 수기록된 하역중단 기록의 적절성을 증진시키는데 활용되었으며, 이와 같이 개선된 하역중단자료는 최종 하역중단 예측모델의 성능을 개선시키는 것으로 연결된다.

예측된 파랑변수와 개선된 하역중단 자료를 이용한 분류모델로 하역중단을 분류할 때, 하역중단이 일어난 경우에 80% 정도를 바르게 분류함으로써 예측모델로서의 상당한 가능성을 보였다. 이와 같이 구성된 분류모델에 새로운 파랑변수 자료를 입력함으로써 얻은 예측된 하역중단은 선박 중량 조건에 따라 하역중단이 일어난 경우 60% 정도 바르게 예측되었다. 하역중단 발생 여부 전체를 바르게 예측한 경우는 80% 이상이었고, 시계열 상으로도 예측모델이 실제 하역중단 경향을 잘 반영하는 것으로 나타났다. 본 하역중단 예측 모델의 성능향상을 위해 더 많은 자료가 사용되고 더 적절한 알고리즘이 적용된다면, 기계학습의 학습 효과에 의해 기존의 방법들로는 얻을 수 없는 상당히 좋은 하역중단 예측 결과를 얻게 될 수 있을 것이다.

keywords: 하역중단, 분류모델, 인공신경망, 힐버트-후앙 트랜스폼, 포항신항

학번: 2005-31040